



Health Technology Assessment

Volume 28 • Issue 43 • August 2024

ISSN 2046-4924

The effects and safety of testosterone replacement therapy for men with hypogonadism: the TestES evidence synthesis and economic evaluation

Moira Cruickshank, Jemma Hudson, Rodolfo Hernández, Magaly Aceves-Martins, Richard Quinton, Katie Gillies, Lorna S Aucott, Charlotte Kennedy, Paul Manson, Nicholas Oliver, Frederick Wu, Siladitya Bhattacharya, Waljit S Dhillon, Channa N Jayasena and Miriam Brazzelli



The effects and safety of testosterone replacement therapy for men with hypogonadism: the TestES evidence synthesis and economic evaluation

Moira Cruickshank¹, Jemma Hudson¹,
Rodolfo Hernández², Magaly Aceves-Martins¹,
Richard Quinton³, Katie Gillies¹, Lorna S Aucott¹,
Charlotte Kennedy², Paul Manson¹, Nicholas Oliver⁴,
Frederick Wu⁵, Siladitya Bhattacharya⁶, Waljit S Dhillon⁴,
Channa N Jayasena⁴ and Miriam Brazzelli^{1*}

¹Health Services Research Unit, University of Aberdeen, Aberdeen, UK

²Health Economics Research Unit, University of Aberdeen, Aberdeen, UK

³Department of Endocrinology, Royal Victoria Infirmary, Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle, UK

⁴Faculty of Medicine, Imperial College London, London, UK

⁵Division of Diabetes, Endocrinology and Gastroenterology, University of Manchester, Manchester, UK

⁶School of Medicine, Medical Sciences and Nutrition, University of Aberdeen, Aberdeen, UK

*Corresponding author

Disclaimer: This report contains transcripts of interviews conducted in the course of the research and contains language that may offend some readers.

Published August 2024
DOI: 10.3310/JRYT3981

This report should be referenced as follows:

Cruickshank M, Hudson J, Hernández R, Aceves-Martins M, Quinton R, Gillies K, *et al*. The effects and safety of testosterone replacement therapy for men with hypogonadism: the TestES evidence synthesis and economic evaluation. *Health Technol Assess* 2024;**28**(43). <https://doi.org/10.3310/JRYT3981>

Health Technology Assessment

ISSN 2046-4924 (Online)

Impact factor: 3.6

A list of Journals Library editors can be found on the [NIHR Journals Library website](#)

Launched in 1997, *Health Technology Assessment* (HTA) has an impact factor of 3.6 and is ranked 32nd (out of 105 titles) in the 'Health Care Sciences & Services' category of the Clarivate 2022 Journal Citation Reports (Science Edition). It is also indexed by MEDLINE, CINAHL (EBSCO Information Services, Ipswich, MA, USA), EMBASE (Elsevier, Amsterdam, the Netherlands), NCBI Bookshelf, DOAJ, Europe PMC, the Cochrane Library (John Wiley & Sons, Inc., Hoboken, NJ, USA), INAHTA, the British Nursing Index (ProQuest LLC, Ann Arbor, MI, USA), Ulrichsweb™ (ProQuest LLC, Ann Arbor, MI, USA) and the Science Citation Index Expanded™ (Clarivate™, Philadelphia, PA, USA).

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: journals.library@nihr.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta.

Criteria for inclusion in the *Health Technology Assessment* journal

Manuscripts are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

Health Technology Assessment (HTA) research is undertaken where some evidence already exists to show that a technology can be effective and this needs to be compared to the current standard intervention to see which works best. Research can evaluate any intervention used in the treatment, prevention or diagnosis of disease, provided the study outcomes lead to findings that have the potential to be of direct benefit to NHS patients. Technologies in this context mean any method used to promote health; prevent and treat disease; and improve rehabilitation or long-term care. They are not confined to new drugs and include any intervention used in the treatment, prevention or diagnosis of disease.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

This article

The research reported in this issue of the journal was funded by the HTA programme as award number 17/68/01. The contractual start date was in September 2018. The draft manuscript began editorial review in September 2021 and was accepted for publication in June 2022. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' manuscript and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this article.

This article presents independent research funded by the National Institute for Health and Care Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, the HTA programme or the Department of Health and Social Care. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, the HTA programme or the Department of Health and Social Care.

This article was published based on current knowledge at the time and date of publication. NIHR is committed to being inclusive and will continually monitor best practice and guidance in relation to terminology and language to ensure that we remain relevant to our stakeholders.

Copyright © 2024 Cruickshank *et al.* This work was produced by Cruickshank *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaptation in any medium and for any purpose provided that it is properly attributed. See: <https://creativecommons.org/licenses/by/4.0/>. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Newgen Digitalworks Pvt Ltd, Chennai, India (www.newgen.co).

Abstract

The effects and safety of testosterone replacement therapy for men with hypogonadism: the TestES evidence synthesis and economic evaluation

Moira Cruickshank¹, Jemma Hudson¹, Rodolfo Hernández²,
Magaly Aceves-Martins¹, Richard Quinton³, Katie Gillies¹,
Lorna S Aucott¹, Charlotte Kennedy², Paul Manson¹,
Nicholas Oliver⁴, Frederick Wu⁵, Siladitya Bhattacharya⁶,
Waljit S Dhillon⁴, Channa N Jayasena⁴ and Miriam Brazzelli^{1*}

¹Health Services Research Unit, University of Aberdeen, Aberdeen, UK

²Health Economics Research Unit, University of Aberdeen, Aberdeen, UK

³Department of Endocrinology, Royal Victoria Infirmary, Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle, UK

⁴Faculty of Medicine, Imperial College London, London, UK

⁵Division of Diabetes, Endocrinology and Gastroenterology, University of Manchester, Manchester, UK

⁶School of Medicine, Medical Sciences and Nutrition, University of Aberdeen, Aberdeen, UK

*Corresponding author m.brazzelli@abdn.ac.uk

Background: Low levels of testosterone cause male hypogonadism, which is associated with sexual dysfunction, tiredness and reduced muscle strength and quality of life. Testosterone replacement therapy is commonly used for ameliorating symptoms of male hypogonadism, but there is uncertainty about the magnitude of its effects and its cardiovascular and cerebrovascular safety.

Aims of the research: The primary aim was to evaluate the safety of testosterone replacement therapy. We also assessed the clinical and cost-effectiveness of testosterone replacement therapy for men with male hypogonadism, and the existing qualitative evidence on men's experience and acceptability of testosterone replacement therapy.

Design: Evidence synthesis and individual participant data meta-analysis of effectiveness and safety, qualitative evidence synthesis and model-based cost-utility analysis.

Data sources: Major electronic databases were searched from 1992 to February 2021 and were restricted to English-language publications.

Methods: We conducted a systematic review with meta-analysis of individual participant data according to current methodological standards. Evidence was considered from placebo-controlled randomised controlled trials assessing the effects of any formulation of testosterone replacement therapy in men with male hypogonadism. Primary outcomes were mortality and cardiovascular and cerebrovascular events. Data were extracted by one reviewer and cross-checked by a second reviewer. The risk of bias was assessed using the Cochrane Risk of Bias tool. We performed one-stage meta-analyses using the acquired individual participant data and two-stage meta-analyses to integrate the individual participant data with data extracted from eligible studies that did not provide individual participant data. A decision-analytic Markov model was developed to evaluate the cost per quality-adjusted life-years of the use of testosterone replacement therapy in cohorts of patients of different starting ages.

Results: We identified 35 trials (5601 randomised participants). Of these, 17 trials (3431 participants) provided individual participant data. There were too few deaths to assess mortality. There was no difference between the testosterone replacement therapy group (120/1601, 7.5%) and placebo group (110/1519, 7.2%) in the incidence of cardiovascular and/or cerebrovascular events (13 studies, odds ratio 1.07, 95% confidence interval 0.81 to 1.42; $p = 0.62$). Testosterone replacement therapy improved quality of life and sexual function in almost all patient subgroups. In the testosterone replacement therapy group, serum testosterone was higher while serum cholesterol, triglycerides, haemoglobin and haematocrit were all lower. We identified several themes from five qualitative studies showing how symptoms of low testosterone affect men's lives and their experience of treatment. The cost-effectiveness of testosterone replacement therapy was dependent on whether uncertain effects on all-cause mortality were included in the model, and on the approach used to estimate the health state utility increment associated with testosterone replacement therapy, which might have been driven by improvements in symptoms such as sexual dysfunction and low mood.

Limitations: A meaningful evaluation of mortality was hampered by the limited number of defined events. Definition and reporting of cardiovascular and cerebrovascular events and methods for testosterone measurement varied across trials.

Conclusions: Our findings do not support a relationship between testosterone replacement therapy and cardiovascular/cerebrovascular events in the short-to-medium term. Testosterone replacement therapy improves sexual function and quality of life without adverse effects on blood pressure, serum lipids or glycaemic markers.

Future work: Rigorous long-term evidence assessing the safety of testosterone replacement therapy and subgroups most benefiting from treatment is needed.

Study registration: The study is registered as PROSPERO CRD42018111005.

Funding: This award was funded by the National Institute for Health and Care Research (NIHR) Health Technology Assessment programme (NIHR award ref: 17/68/01) and is published in full in *Health Technology Assessment*; Vol. 28, No. 43. See the NIHR Funding and Awards website for further award information.

Contents

List of tables	xi
List of figures	xiii
List of abbreviations	xvii
Plain language summary	xix
Scientific summary	xxi
Chapter 1 Background and research question	1
Description of underlying health problem	1
Epidemiology and prevalence	1
Impact of health problem	1
Guidelines for measurement, diagnosis and treatment of condition	1
Current usage of testosterone replacement therapy in the NHS	2
Decision problem	2
Description of interventions under assessment	2
Population and relevant subgroups	3
Overall aim and objectives of this assessment	3
Chapter 2 Quantitative synthesis and individual participant data meta-analysis	5
Methods for assessing the outcomes arising from the use of the intervention	5
<i>Protocol and registration</i>	5
<i>Eligibility criteria</i>	5
<i>Identifying studies: information sources and search strategy</i>	7
<i>Study selection process</i>	8
<i>Data collection processes</i>	8
<i>Data items and individual participant data integrity</i>	9
<i>Risk of bias assessment in individual studies</i>	9
<i>Statistical analyses</i>	9
Results of the quantitative synthesis and individual participant data meta-analysis	11
<i>Study selection and individual participant data obtained</i>	11
<i>Study characteristics</i>	12
<i>Participant characteristics</i>	16
<i>Individual participant data integrity</i>	22
<i>Risk of bias within studies</i>	23
<i>Primary outcomes</i>	27
<i>Secondary outcomes</i>	29
Discussion	37
Conclusions	44
Chapter 3 Synthesis of qualitative and mixed-methods evidence evaluating men's experiences of low testosterone and/or health professionals' and care providers' views of testosterone replacement therapy	45
Introduction	45
<i>Value of mixed-methods qualitative studies</i>	45
<i>Role for qualitative studies in male hypogonadism</i>	45

CONTENTS

Methods	45
<i>Searching and identification of relevant studies</i>	45
<i>Data extraction and synthesis</i>	46
<i>Qualitative analysis</i>	46
<i>Quality assessment strategy</i>	47
<i>Assessment of confidence in the review findings</i>	47
Findings	47
<i>Description of included studies</i>	47
<i>Overall findings</i>	51
<i>Quality assessment results</i>	62
<i>Confidence in the findings</i>	62
Discussion	67
<i>Limitations</i>	69
<i>Conclusions</i>	69
Chapter 4 Disease-specific patient-reported outcome measures for low testosterone: an investigation of item content to assess conceptual comparability	71
Introduction	71
Methods	71
<i>Search methods for identification of studies</i>	71
<i>Inclusion criteria</i>	72
<i>Exclusion criteria</i>	72
<i>Study selection and data extraction</i>	72
<i>Data synthesis</i>	73
Results	73
<i>Descriptive characteristics: included studies</i>	73
<i>Descriptive characteristics: PROMs from included studies</i>	73
<i>Item domain classification</i>	81
Discussion	83
<i>Strengths and limitations</i>	84
<i>Conclusions</i>	84
Chapter 5 Cost-effectiveness of testosterone replacement therapy	85
Systematic review of existing cost-effectiveness	85
<i>Inclusion and exclusion criteria</i>	85
<i>Search strategy</i>	85
<i>Study selection and data extraction</i>	85
<i>Results of the systematic review of cost-effectiveness</i>	85
Model-based economic evaluation	86
<i>Introduction</i>	86
<i>Methods</i>	86
<i>Results</i>	97
<i>Discussion</i>	107
<i>Conclusions</i>	109
Chapter 6 Conclusions	111
Implications for health care	111
<i>Recommendations for future research</i>	112
<i>Patient and public involvement</i>	113

Additional information	115
References	119
Appendix 1 Ovid (MEDLINE and Embase) search strategy	129
Appendix 2 Data extraction form	135
Appendix 3 List of key items requested from authors of existing trials	139
Appendix 4 Standard operating procedure for management of IPD	143
Appendix 5 Study characteristics and participant characteristics of IPD and non-IPD studies	145
Appendix 6 Clinical effectiveness results	169
Appendix 7 Details of the health-related quality-of-life data and analysis for the economic model	209

List of tables

TABLE 1	Summary of the characteristics of the 17 studies that provided IPD	13
TABLE 2	Summary of the characteristics of the 18 studies that did not provide IPD	14
TABLE 3	Summary of the characteristics of participants enrolled in the 17 studies that provided IPD	17
TABLE 4	Summary of the characteristics of participants enrolled in the 18 studies that did not provide IPD	18
TABLE 5	Baseline characteristics of the participants enrolled in the 17 studies that contributed to the IPD analysis	20
TABLE 6	One-stage IPD meta-analysis for mortality from any cause	29
TABLE 7	One-stage IPD meta-analysis for CV and/or CBV events	31
TABLE 8	One-stage analysis for QoL outcomes	33
TABLE 9	One-stage analysis for sexual function outcomes	36
TABLE 10	One-stage meta-analysis for physiological marker outcomes	40
TABLE 11	Summary of other outcomes	43
TABLE 12	Participant characteristics of included studies	48
TABLE 13	Summary of the key concepts identified in literature	52
TABLE 14	Contributions of included studies to themes and subthemes	53
TABLE 15	CERQual evidence profile	63
TABLE 16	Main characteristics of PROM studies	74
TABLE 17	Characteristics of included PROMs	79
TABLE 18	Domain definitions	81
TABLE 19	Domains identified across relevant individual items from the assessed PROMs	82
TABLE 20	IPD meta-analysis outcomes incorporated into the economic model	88
TABLE 21	Proportion of CV and CBV complications	88
TABLE 22	Underlying monthly risk of CV events	89
TABLE 23	Utility scores for the <i>No Complication</i> health state for TRT and SoC	90

TABLE 24	Monitoring test schedule. Unit costs for tests and monitoring visits	92
TABLE 25	Utility multipliers attached to CV complications	93
TABLE 26	Unit costs attached to CV complications	95
TABLE 27	Summary of the economic evaluation scenarios	97
TABLE 28	Scenario analysis results. Model starting age 40 years	98
TABLE 29	Scenario analysis results. Model starting age 60 years	100
TABLE 30	Scenario analysis results. Model starting age 75 years	102
TABLE 31	Details of items collected from the included studies	136
TABLE 32	List of key items requested from trials' authors	139
TABLE 33	Study characteristics of the 17 studies that provided IPD	146
TABLE 34	Study characteristics of the 18 studies that did not provide IPD	152
TABLE 35	Participant characteristics of the 17 studies that provided IPD	159
TABLE 36	Participant characteristics of the 18 studies that did not provide IPD	163
TABLE 37	Baseline characteristics of patients in studies that contributed to the IPD	169
TABLE 38	Two-stage IPD meta-analysis for mortality from any cause	171
TABLE 39	Two-stage IPD meta-analysis for CV and/or CBV events	171
TABLE 40	One-stage analysis for QoL outcomes	178
TABLE 41	One-stage analysis for sexual function outcomes	179
TABLE 42	Subgroup analysis for IIEF-15 and its subscales	185
TABLE 43	Sexual function during testosterone treatment in men classified according to identified thresholds	186
TABLE 44	One-stage analysis for physiological marker outcomes	188
TABLE 45	One-stage meta-analysis analysis for psychological symptom outcomes	205
TABLE 46	SF-36 data available from received studies	209
TABLE 47	Utility score difference between TRT and no TRT for the SF-6D data analysis	209
TABLE 48	Beck Depression Inventory score (BDI) data available from received studies	210
TABLE 49	Utility score difference between TRT and no TRT for the EQO-5D score from the mapping exercise	210

List of figures

FIGURE 1 PRISMA-IPD flow diagram of selected studies	12
FIGURE 2 Summary of the risk of bias assessment of the 17 IPD studies (data derived from published reports and from contacting the collaborators)	23
FIGURE 3 Risk of bias assessments of the 17 IPD studies (data derived from published reports and from contacting the collaborators)	24
FIGURE 4 Summary of the risk of bias assessment of the 17 IPD studies (data from published reports only)	25
FIGURE 5 Risk of bias assessments of the 17 IPD studies (data from publications only)	26
FIGURE 6 Summary of risk of bias assessment of the 18 non-IPD studies	27
FIGURE 7 Risk of bias assessments of the 18 non-IPD studies	28
FIGURE 8 IPD meta-analysis for mortality from any cause	30
FIGURE 9 Two-stage IPD meta-analysis for CV and/or CBV events	32
FIGURE 10 Subgroups for TRT vs. placebo	32
FIGURE 11 Two-stage meta-analysis for AMS	34
FIGURE 12 Two-stage meta-analysis for IIEF-15	37
FIGURE 13 Subgroups for TRT vs. placebo	41
FIGURE 14 PRISMA flow diagram of qualitative study selection	46
FIGURE 15 Conceptual diagram of the evidence on men and their healthcare providers' experiences of low testosterone and TRT.	67
FIGURE 16 PRISMA flow diagram for identification of the PROM studies.	74
FIGURE 17 Simplified schematic of the economic model structure.	87
FIGURE 18 Cost-effectiveness acceptability curves for TRT for alternative scenarios – 40-year-old cohort.	105
FIGURE 19 Cost-effectiveness acceptability curves for TRT for alternative scenarios – 60-year-old cohort.	106
FIGURE 20 Cost-effectiveness acceptability curves for TRT for alternative scenarios – 75-year-old cohort.	106
FIGURE 21 Contour-enhanced funnel plots for mortality from any cause	172

FIGURE 22	Contour-enhanced funnel plots for CV and/or CBV events	173
FIGURE 23	Two-stage meta-analysis for SF-36/SF-12 norm-based scores	173
FIGURE 24	Two-stage meta-analysis for IIEF-5.	179
FIGURE 25	Two-stage meta-analysis for HED baseline	180
FIGURE 26	Two-stage meta-analysis for Sexual Arousal, Interest, and Drive Scale	180
FIGURE 27	Two-stage meta-analysis for PDQ	180
FIGURE 28	Two-stage meta-analysis for Male Sexual Health Questionnaire-Ejaculatory Dysfunction-Short Form	184
FIGURE 29	Thresholds for sexual function experienced during testosterone treatment in men	186
FIGURE 30	Scatterplot of baseline IIEf-15 vs. serum baseline TT	187
FIGURE 31	Two-stage meta-analysis for increase in testosterone (nmol/l)	189
FIGURE 32	Two-stage meta-analysis for increase in free testosterone (pmol/l)	190
FIGURE 33	Two-stage meta-analysis for increase in fasting glucose (mmol/l)	190
FIGURE 34	Two-stage meta-analysis for increase in cholesterol (mmol/l)	191
FIGURE 35	Two-stage meta-analysis for increase in low-density lipoprotein cholesterol (mmol/l)	192
FIGURE 36	Two-stage meta-analysis for increase in high-density lipoprotein cholesterol (mmol/l)	193
FIGURE 37	Two-stage meta-analysis for increase in triglycerides (mmol/l)	194
FIGURE 38	Two-stage meta-analysis for increase in Hb (g/l)	194
FIGURE 39	Two-stage meta-analysis for increase in HbA1c (%)	195
FIGURE 40	Two-stage meta-analysis for increase in haematocrit (%)	196
FIGURE 41	Two-stage meta-analysis for increase in SBP (mmHg)	197
FIGURE 42	Two-stage meta-analysis for increase in DBP (mmHg) (nmol/l)	198
FIGURE 43	Two-stage meta-analysis for increase in bone mineral density (g/cm ²)	198
FIGURE 44	Two-stage meta-analysis for BDI	206

FIGURE 45 Two-stage meta-analysis for the Geriatric Depression Scale	206
FIGURE 46 Two-stage meta-analysis for Hamilton Depression Scale	207
FIGURE 47 Two-stage meta-analysis for Profile of Mood States	207

List of abbreviations

ADAM	androgen deficiency in ageing males	HRG	Healthcare Resource Group
AMS	ageing males' symptoms	HRQoL	health-related quality of life
A-RHDQoL	age-related hormone deficiency dependent quality of life questionnaire	ICER	incremental cost-effectiveness ratio
BDI	Beck depression inventory	IIEF	International Index of Erectile Function
BMI	body mass index	IIEF-5	International Index of Erectile Function-5 items
BNF	British National Formulary	IIEF-15	International Index of Erectile Function-15 items
CASP	Critical Appraisal Skills Programme	IPD	individual participant data
CBV	cerebrovascular	IQA	internal quality control
CEAC	cost-effectiveness acceptability curves	LDL	low-density lipoprotein
CHD	coronary heart disease	HDL	high-density lipoprotein
CI	confidence interval	LFT	liver function test
CV	cardiovascular	LOH	Late-onset hypogonadism
DBP	diastolic blood pressure	MACE	major adverse cardiovascular events
DISF-II	Derogatis Interview for Sexual Functioning in men-II5	MD	mean difference
EMAS	European Male Ageing Study	MH	male hypogonadism
EQ-5D	EuroQol-5 dimensions	MMAS	Aging Study
EQA	external quality control	NICE	National Institute for Health and Care Excellence
GP	general practitioner	NIHR	National Institute for Health and Care Research
GRADE-CERQual	grading of recommendations assessment, development and evaluation-confidence in the evidence from reviews of qualitative research	OR	odds ratio
HADS	Hospital Anxiety and Depression Scale	PDQ	psychosexual daily questionnaire
Hb	haemoglobin	PROM	patient-reported outcome measure
HbA1c	glycated haemoglobin	PSA	prostate-specific antigen
HED	hypogonadism energy diary	PSSRU	Personal Social Services Research Unit
HIS-Q	hypogonadism impact of symptoms questionnaire	QALY	quality-adjusted life-year
HIS-Q-SF	hypogonadism impact of symptoms questionnaire-short form	QoL	quality of life
		RCT	randomised controlled trial
		REML	restricted maximum likelihood
		RR	relative risk

LIST OF ABBREVIATIONS

SAID	sexual arousal, interest and drive scale	T2D	type 2 diabetes
SBP	systolic blood pressure	T2DM	type-2 diabetes mellitus
SD	standard deviation	TRT	testosterone replacement therapy
SF-6D	short form-6 dimensions	TT	total testosterone
SF-12	short form-12 items	TU	testosterone undecanoate
SF-36	short form-36 items	WHO-ICF	World Health Organization International Classification of Functioning, Disability and Health
SMR	standardised mortality ratio		
SoC	standard of care		
SOP	standard operating procedure		

Plain language summary

Testosterone is a hormone which is vital for sexual activity, bone growth and muscle development in men. Men with low testosterone levels may experience problems with erections and may suffer from brittle bones (osteoporosis), weakness, feeling down (low mood) and tiredness. The manifestations of low testosterone can be treated with testosterone replacement therapy. However, there is current uncertainty about the positive effects of testosterone replacement therapy and its safety.

We brought together results from all available medical studies that looked at the use of testosterone replacement therapy in men with low testosterone and contacted the doctors who led these studies to gather further information on their participants.

We found 35 studies (5601 participants) conducted in different countries, 17 of which provided additional information on their participants. We did not find any evidence to show that testosterone replacement therapy increases the risk of heart problems, or any evidence to show that some men who take testosterone replacement therapy benefit more than others. Men with low testosterone reported having low mood, poor concentration and lack of energy; however, medical studies often failed to prove that these manifestations improved with testosterone replacement therapy. Most medical studies were conducted among white men in North America using questionnaires designed specifically for them; therefore, the results may not reflect the experiences of men in other countries and from more diverse ethnic backgrounds. There is too much uncertainty about the benefits of testosterone replacement therapy to accurately estimate its value for money for the NHS.

We think our findings offer some reassurance to doctors and patients that testosterone replacement therapy does not increase the risk of heart problems. New studies are needed to find out whether some groups of men (such as older or younger men) are more likely to benefit from testosterone replacement therapy more than others. It is also important to develop tools which better reflect the experience of men from a diverse range of social and ethnic backgrounds. To inform men with low testosterone about our findings, we are creating a website with dedicated YouTube video clips.

Scientific summary

Background and objectives

Low levels of the testicular androgen hormone testosterone cause male hypogonadism (MH), which is associated with sexual dysfunction, hot flushes, reduced physical energy, mood disturbance and complications such as osteoporosis. Testosterone (androgen) replacement therapy (TRT) is the standard of treatment for MH. Prescribing rates of TRT in the UK have risen between 2001 and 2010 without an observable increase in the prevalence of hypogonadism, which may reflect increasing media awareness among men about MH. Randomised controlled trials (RCTs) of TRT in men with MH have produced varying and sometimes conflicting results regarding the safety and efficacy of TRT. In particular, there is considerable uncertainty regarding the impact of TRT on cardiovascular (CV) event risk in men with MH, which has led to mandatory safety labelling in some countries. Lack of clarity about the effects of TRT have led to variations in clinical practice among clinicians, which in turn exposes men with MH to inconsistent standards of clinical care and potentially life-threatening healthcare risks. The overarching aim of this project was to conduct a comprehensive and unbiased appraisal of quantitative, qualitative and economic evidence for the use of TRT in symptomatic men with MH. This evidence synthesis had three specific objectives: (1) to perform an individual patient data meta-analysis to estimate the clinical effectiveness and safety of TRT for men with testosterone deficiency syndrome; (2) to synthesise the existing qualitative evidence reporting men's experience and acceptability of TRT; (3) to develop a decision model to estimate the cost-effectiveness of TRT for the treatment of symptomatic men with testosterone deficiency syndrome.

Review methods

Methods for the quantitative synthesis and individual participant data meta-analysis

We conducted a quantitative synthesis including a meta-analysis of individual participant data (IPD) according to current methodological standards, to identify evidence from placebo-controlled RCTs evaluating the effects of TRT in men with low testosterone. The methods were pre-specified in a research protocol (PROSPERO database registration number: CRD42018111005; www.crd.york.ac.uk/prospero/display_record.php?RecordID=111005).

Comprehensive searches of major electronic databases were conducted, including MEDLINE, MEDLINE In-process & Other Non-indexed Citations, MEDLINE Epub Ahead of Print, EMBASE, Science Citation Index and CENTRAL. Searches were restricted to publications from 1992 to date. Searches were carried out in August 2018 and updated in February 2021. The population considered was men with low testosterone. The intervention was TRT and the comparator was placebo. Primary outcomes were all-cause mortality and CV and cerebrovascular (CBV) events. Other outcomes of interest were sexual function, physical health parameters, functional activities, psychological symptoms, other comorbidities, prostate-related outcomes, physiological markers and quality of life (QoL). Outcomes were assessed at 12 months or at the closest time point to 12 months. Two reviewers independently screened all records identified by the searches and one reviewer screened all potentially relevant full texts, with a 10% sample independently checked by a second reviewer. Data were extracted by one reviewer with a random sample of 10% cross-checked by a second reviewer and all outcome data further cross-checked by the statistician. Risk of bias of the included RCTs was assessed by two independent reviewers using the Cochrane Risk of Bias tool (version 1). Data were analysed using both a one-stage meta-analysis for each outcome using the acquired IPD and a two-stage approach to enable the integration of the IPD along with the extracted published data from eligible studies without IPD.

Methods for qualitative synthesis and patient-reported outcome measures analysis

A comprehensive search was conducted to identify published papers reporting qualitative data and/or the development of patient-reported outcome measures (PROMs) for men with hypogonadism who received or were considered to receive TRT, and papers reporting the views of care providers. Electronic databases (Ovid MEDLINE, Embase, PsycInfo, EBSCO CINAHL and Proquest ASSIA) were searched for publication from 1992 to February 2020. The same eligibility criteria used for the IPD meta-analysis were applied. One review author (MA-M) independently screened all titles and abstracts with a randomly selected sample of 10% cross-checked by a second review author (KG). A third author (JC) was consulted when consensus could not be reached regarding eligibility.

For the qualitative evidence synthesis, we appraised eligible studies for methodological rigour and theoretical relevance using the Critical Appraisal Skills Programme (CASP) tool. We conducted a thematic synthesis using both inductive and deductive approaches to analysis and we applied grading of recommendations assessment, development and evaluation-confidence in the evidence from reviews of qualitative research (GRADE-CERQual) to the findings of the thematic synthesis.

For the PROMs analysis, we analysed the individual verbatim items from each PROM using a directed content analysis approach informed by the included papers and the relevant health domain was the World Health Organization International Classification of Functioning, Disability and Health (WHO-ICF). Domain mapping was conducted by two reviewers (MA-M and KG) independently, with any conflicts resolved through discussion. Descriptive statistics were used to describe general information and measure detail with narrative synthesis of the PROM and their inter-related domains.

Methods for the systematic review of economic evaluations and the development of a new economic model

The economic evidence on TRT was assessed through a systematic review (SR) of economic evaluations and a new model-based economic evaluation comparing TRT with standard of care (SoC; e.g. no treatment).

Systematic review of economic evaluations

Full economic evaluations (i.e. studies reporting cost and consequences – regardless of the way these were estimated – for at least two strategies including TRT) were included. Relevant electronic databases [i.e. MEDLINE, Embase, NHS Economic Evaluations Database (NEED), the HTA Database, Cost-Effectiveness Analysis Registry and Research Papers in Economics] were searched from 1992 until 4 February 2021. In addition, conference proceedings of key professional organisations (endocrinology, cardiology and men’s health) from 2018 to 2020 and International Society for Pharmacoeconomics and Outcomes Research (ISPOR) Scientific Presentations Databases were searched. One reviewer screened titles and abstracts, selected studies for inclusion and extracted data following the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) checklist.

Economic model

A cost–utility analysis was conducted using a cohort Markov model that was developed following best practice on decision modelling. The model care pathways were informed by existing clinical guidelines, the results of the IPD meta-analysis, and by discussion with clinical and methodological experts. Similar structures including five Markov states (no complications, post-Cardiac pathology, post-Peripheral Vascular System Pathology, post-Cerebrovascular System Pathology and Death) were used for TRT and SoC. The analysis of the TestES IPD provided key input data (i.e. all-cause mortality, CV and CBV complications and utility weights for TRT and SoC). NHS and personal and social services perspectives were adopted for costs. Ten-year and lifetime time horizons were considered. Cost and effects were discounted at 3.5% annual discount rate. We attached probability distributions to input mean parameter values and the model was run probabilistically using 10,000 Monte Carlo simulations. Three cohorts were defined according to the patients’ starting age (40-year-old cohort, 60-year-old cohort, 75-year-old cohort). Mean total cost, mean quality-adjusted life-years (QALYs) and incremental

cost-effectiveness ratios (ICERs) were calculated. Probabilistic results were reported using cost-effectiveness acceptability curves (CEAC).

Results

Results of the quantitative synthesis and individual participant data meta-analysis

Thirty-five trials with a total of 5601 randomised participants were included in the review of clinical effectiveness. Of these, 17 studies provided IPD for a total of 3431 participants. Overall risk of bias was assessed as being low for 12/17 studies providing IPD and unclear for the remaining five studies. Of the 18 studies not providing IPD, risk of bias was assessed as low for three studies, unclear for 13 studies and high for two studies. Mortality was lower in the TRT group (0.4%) than in the placebo group (0.8%) [odds ratio (OR) 0.46, 95% confidence interval (CI) 0.17 to 1.24; $p = 0.13$], but there were too few events for a reliable evaluation. There was no difference in the occurrence of CV and/or CBV events in the TRT group (7.5%) and the placebo group (7.2%; OR 1.07, 95% CI 0.81 to 1.42; $p = 0.62$). Effects of TRT on QoL were more variable, with some scales showing a difference in favour of TRT [International Index of Erectile Function-15 items (IIEF-15), ageing males' symptoms (AMS), 3/10 subscales of the short form-36 items/short form-12 items (SF-36/SF-12)]. Serum testosterone was higher in the TRT group and serum cholesterol, triglycerides, haemoglobin and haematocrit were all lower in the TRT than the placebo group.

Results of the qualitative synthesis and patient-reported outcome measures

Five studies, reported in nine publications, were included in the qualitative evidence synthesis. Six broad themes (with several linked subthemes) were identified in relation to men's, and their care providers', experiences of low testosterone and receiving TRT as treatment. Five broad patient-facing themes were identified and ordered to reflect key timeline stages and decision points that a man with low testosterone may experience: symptoms of low testosterone and impact on daily life; diagnosis of low testosterone; access to treatment information; perceived effects of TRT; expectations, experience and preference of the type of TRT. A sixth theme on providers' perception of diagnosis and treatment was also identified. We identified several interconnected subthemes highlighting the complexity of how low testosterone symptoms influence many aspects of men's lives and their experiences of treatment.

The PROMs search identified a total of nine eligible PROMs measuring experiences of men with low testosterone. The number of items varied across PROMs and ranged from 3 to 53 items (median = 7) with a cumulative total of 98 individual items across the nine PROMs. Our review identified 10 relevant health domains across the 98 items from the 9 PROMs. The 10 domains were defined according to the WHO-ICF classifications and were identified as: Cognition, Energy, General Well-being, Mood, Pain, Physical-General, Role, Sexual, Sleep, Social. The domain most frequently identified across PROMs was the sexual domain, with 29 (29.6% of total items) items measuring this concept across the PROMs.

Results of the systematic review of economic evaluations and of the cost-effectiveness analyses

Only one study met our inclusion criteria for the systematic review of economic evaluations. The study was a model-based cost-utility analysis conducted in Sweden. The model accounted for TRT benefits from reduced risk of fractures, depression, and type 2 diabetes and for a higher risk of CV and CBV events associated with diabetes. The authors concluded that the lifelong treatment with testosterone undecanoate (TU) depot injection was a cost-effective treatment option for men diagnosed with hypogonadism in Sweden.

Our model results show that the cost-effectiveness of TRT is dependent on the relative risk (RR) of all-cause mortality and the methods used to derive health state utility scores [i.e. through the short form-6 dimensions (SF-6D) algorithm or a mapping exercise between Beck depression inventory (BDI) score and EuroQoL-5 dimensions (EQ-5D) score] for the TRT versus SoC. When the RR of mortality favouring TRT

and the BDI-based utility scores were used for the 10-year time horizon, ICERs remained below the £20,000 threshold, irrespective of the cohort starting age. ICERs also stayed below the £20,000 threshold for the 60- and 75-year-old cohorts when the RR of mortality favouring TRT and the SF-6D utility difference were applied for 10 years. However, ICERs increased above the £20,000 threshold when the difference in all-cause mortality between TRT and SoC was dropped, and the utility scores were defined using the SF-6D. Extending the model time horizon for lifetime for the later scenario further increased the ICER as the impact of complications became more pronounced, eroding the modest SF-6D-based QALY increment.

Limitations

- Inclusion was restricted to studies published in English.
- Lack of long-term data on the effects and safety of TRT.
- A meaningful evaluation of the association between mortality and TRT was hampered by the limited number of defined events.
- Definition and reporting of CV and CBV events and methods for testosterone measurement varied across studies.
- Length of follow-up in many included trials precluded the accumulation of enough CV or CBV events.
- There were too few studies of oral testosterone to assess whether the incidence of CV or CBV events was affected by the mode of administration of testosterone therapy.
- There were only a limited number of available qualitative studies assessing the experience of men with low testosterone; these were predominantly conducted in North America among white men.
- There were limited data for generating preference-based health-related QoL weights to be used in the economic evaluation.

Implications for health care

Our collaborative IPD meta-analysis failed to show a relationship between TRT and CV or CBV events in the short-to-medium term (within 12 months). Our findings indicate that TRT improves sexual function and patients' QoL without adverse effects on blood pressure, serum lipids or glycaemic markers.

Overall, there is a paucity of qualitative evidence on the effects of low testosterone and consequences of TRT when compared with the number of existing clinical trials assessing the effectiveness of TRT. Our results indicate that the effects of low testosterone and subsequent treatment with TRT have multiple impacts and concerns for men. This study has shown the considerable variability that exists in disease-specific PROMs for men with low testosterone regarding development and domain coverage. It has also highlighted the lack of input from men in the development of these PROMs, bringing into questions their relevance and adequacy in capturing outcomes that matter to men with low testosterone.

Results of our economic model suggest that the cost-effectiveness of TRT is dependent on its effects on all-cause mortality and health state utility resulting from improvements in symptoms associated with hypogonadism such as sexual dysfunction, low mood and reduced QoL. We also noticed that the choice of the instrument and approach to estimate QoL weights (BDI or SF-6D) was crucial to the cost-effectiveness of TRT. Further clarity on the CV safety of TRT in men with hypogonadism and more in-depth mapping of clinical outcomes to generic preference-based measures of health-related QoL will be crucial to inform more robust estimates of the cost-effectiveness of TRT for men with hypogonadism.

Recommendations for future research

Future research should assess the safety of long-term use of TRT including its impact on mortality, identify the threshold of baseline serum testosterone associated with improvement of sexual function, cognition, mood and QoL in men with hypogonadism and compare the effects of transdermal versus intramuscular TRT. Moreover, further research aiming at mapping QoL instruments commonly used in existing trials of MH to generic preference-based health-related QoL instruments would facilitate the integration of existing evidence into future economic evaluations and models. There is also the need to gather more data on the experience of men with hypogonadism from multiethnic populations and develop a holistic symptom score for TRT response.

Study registration

The study is registered as PROSPERO CRD42018111005.

Funding

This award was funded by the National Institute for Health and Care Research (NIHR) Health Technology Assessment programme (NIHR award ref: 17/68/01) and is published in full in *Health Technology Assessment*; Vol. 28, No. 43. See the NIHR Funding and Awards website for further award information.

Chapter 1 Background and research question

Description of underlying health problem

Testosterone is the major male sex hormone (androgen) produced in the testes. Testosterone is essential for normal sexual function, muscle growth, haematopoiesis and bone mineralisation, and has important behavioural effects in men. Male hypogonadism (MH) is a clinical syndrome of low testosterone associated with symptoms such as sexual dysfunction, hot flushes, reduced physical energy and mood disturbance, and complications such as osteoporosis, gynaecomastia or anaemia.¹ However, many of the associated symptoms are non-specific and may be caused by comorbidities such as obesity and depression rather than low testosterone itself. Furthermore, low testosterone may result from obesity and type 2 diabetes, making it difficult to interpret observational studies of their associations.¹ MH may be caused by testicular or pituitary disease, either extreme of body fat, diabetes, ageing, drugs or genetic conditions.² Testosterone replacement therapy (TRT) is the accepted standard of treatment in men with MH.³

Epidemiology and prevalence

Levels of circulating testosterone decline annually by approximately 1% from the age of 40 years onwards and 33%–50% of middle-aged men have decreased serum levels of the hormone.^{4–6} Only a minority of men with low testosterone levels fulfil the criteria for MH. The European Male Ageing Study (EMAS) concluded that approximately 2% of over 3400 men aged 40–79 years had MH.⁴ There is no universally accepted definition of MH, and so its estimated prevalence varies according to the stringency of the diagnostic criteria.^{7,8}

Impact of health problem

Symptomatic MH results in significant morbidity (erectile dysfunction, low mood, osteoporosis, muscle weakness, anaemia, gynaecomastia), impairing quality of life (QoL), cognition, mental health and daily function, and ultimately impacts on men's ability to live well for longer.⁹

Guidelines for measurement, diagnosis and treatment of condition

Male hypogonadism is defined by the combination of characteristic clinical features and corroborative biochemistry.⁷ However, there is currently no unifying consensus for the biochemical criteria for diagnosing MH. One approach is to calculate a statistical reference range based on healthy, non-obese young men without sexual dysfunction and excluding those with raised gonadotropin levels; the United States (US) Centers for Disease Control and Prevention (CDC) determined in 9000 men that the 2.5th and 97.5th percentiles were 9.2 nmol/l and 31.8 nmol/l, respectively.¹⁰ The EMAS recruited over 3000 men aged 40–79 years across Europe; the probability of experiencing sexual dysfunction increased if serum total testosterone (TT) was either < 8 nmol/l or < 11 nmol/l (with calculated free testosterone < 220 pmol/l).⁴ The National Institute of Health (NIH) utilised TT thresholds of 9.5 and 10.4 nmol/l for inclusion of older men to their randomised controlled trial (RCT) studies of TRT.¹¹ However, a large RCT that was suspended early due to unexpected cardiovascular (CV) events used a higher TT threshold of 12.1 nmol/l.¹² Other RCTs have included men with TT levels up to 14 nmol/l. Considering published clinical guidelines for diagnosing MH, several possible criteria have been proposed, including serum TT diagnostic thresholds ranging from 8 to 12 nmol/l,⁸ however, none of the guidelines propose

that a serum TT above 12 nmol/l is routinely consistent with the diagnosis of MH. TRT injections or transdermal gels have been shown during RCTs to be highly effective at relieving symptoms of hypogonadism, and are licensed accordingly for treating affected men.¹¹

Current usage of testosterone replacement therapy in the NHS

Testosterone replacement therapy is an established therapy for MH. As MH is more common in middle- and older-aged men, use of TRT is likely to increase in the future. A recent UK study observed that NHS prescriptions of androgen replacement therapy (TRT, commonly known as ‘testosterone therapy’) for men had doubled since 2001, at an increased annual cost of £8 million; however, the incidence of low testosterone was unchanged over the 10-year time frame of this study.¹³ This raises the possibility that increased media and societal awareness of testosterone as treatment for ‘andropause’ may have shifted the views of some patients and/or clinicians to lower the threshold for TRT prescribing.^{14,15}

Decision problem

A small number of clinical studies have suggested that TRT may increase the risk of CV disease, which has triggered safety concerns among some patients and physicians.¹⁶ Nevertheless, existing systematic reviews of randomised trials present conflicting results regarding the CV safety and clinical effects of TRT.¹⁷⁻²⁶ This lack of robust evidence on the effects and safety of TRT in men has polarised opinion among the clinical and scientific communities regarding the treatment of MH; consequently, clinical guidelines and definitions of MH vary considerably.^{7,8} This, in turn, exposes men to inconsistent standards of clinical care, and potentially life-threatening healthcare risks (whether arising from necessary TRT withheld, or unnecessary TRT prescribed). Several meta-analyses of published RCT data have respectively concluded that TRT increases CV risk, reduces CV risk or has no significant effect on CV risk in men with MH.¹⁷⁻²⁶ This unsatisfactory situation may partially reflect a lack of consistent adverse CV event classification or reporting within RCTs. Determining the safety together with the clinical and cost-effectiveness of TRT in men with MH is critical to inform decision-making by men, their clinicians, healthcare providers and policy-makers.

Previous studies have recruited heterogeneous populations and have used a variety of clinical tools measuring different aspects of illness and outcomes, making it difficult for clinicians to compare results across studies. A limited number of qualitative studies have also explored the perceptions of men on TRT, but these data have not been systematically synthesised hitherto.²⁷⁻²⁹ Understanding men’s expectations and experiences of TRT and the influence this has on their QoL would contribute further evidence towards determining for whom this intervention may be most relevant and beneficial. It is also possible that the symptomatic effects of TRT are dependent on factors including patient age, serum testosterone levels and pre-existing comorbidities (e.g. depression, type 2 diabetes, poor mobility).¹ In summary, TRT offers several potential symptomatic benefits for men with MH, but suffers from a conflicting array of evidence which compromises clinical decision-making across the NHS.

Description of interventions under assessment

Testosterone replacement therapy with any formulation, dose, frequency and route of administration. Studies that use other androgens apart from testosterone and studies allowing concurrent treatment with other hormones or interventions will be excluded.

Population and relevant subgroups

Men with MH, that is, symptoms suggestive of low testosterone and low levels of serum testosterone. Studies restricted to specific, syndromic forms of MH such as congenital hypogonadotropic hypogonadism (CHH), hypopituitarism or Klinefelter syndrome are outside the remit of this evidence synthesis, which was to focus on the effects of MH per se. As discussed in the section on clinical guidelines, a serum TT threshold < 12 nmol/l was used for including published data.

Overall aim and objectives of this assessment

The overarching aim is to review the existing quantitative, qualitative and economic evidence for the use of TRT monotherapy in symptomatic men with MH.

Specific objectives:

- To conduct an evidence synthesis including an individual participant data (IPD) meta-analysis to estimate the clinical effectiveness and safety of TRT for men with testosterone deficiency syndrome.
- To synthesise the existing qualitative evidence and patient-reported outcome measures (PROMs) which report men's experience and acceptability of TRT.
- To develop a decision model to estimate the cost-effectiveness of TRT for the treatment of symptomatic men with testosterone deficiency syndrome.

Chapter 2 Quantitative synthesis and individual participant data meta-analysis

We conducted an objective synthesis of the current evidence assessing the clinical effectiveness and safety of testosterone for men with low testosterone. The evidence synthesis was carried out according to the general principles of the Centre for Reviews and Dissemination (CRD) guidance for undertaking reviews in health care, and the recommendations of the Cochrane Handbook for Systematic Reviews of Interventions, and was reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses – Individual Participant Data (PRISMA-IPD) checklist.

Methods for assessing the outcomes arising from the use of the intervention

Protocol and registration

The methods were pre-specified in a research protocol (PROSPERO database registration number: CRD42018111005; www.crd.york.ac.uk/prospero/display_record.php?RecordID111005).

Eligibility criteria

Study design

Evidence was considered from randomised placebo-controlled clinical trials evaluating the effects of TRT in men with low testosterone. Only trials with a duration of at least 3 months for all intervention groups were considered suitable for inclusion; this was in line with the current recommendation of the UK British Society for Sexual Medicine (which recommends evaluating patients at 3, 6 and 12 months after TRT initiation and then every 12 months) and US Endocrine Society Clinical Practice Guidelines (which recommend evaluating men 3–6 months after TRT initiation and then annually thereafter).^{30,31} Any relevant clinical setting (e.g. primary care, secondary care) was eligible for inclusion. Studies of a cross-over design were not considered suitable for inclusion, unless there were eligible groups with sufficient follow-up before cross-over occurred.

Target population

Adult men (aged 18 years or over with no upper age limit) presenting with a proven low level of serum testosterone. There has never been a consensus definition of low testosterone, which is reflected by the participant characteristics of trials in this field. However, all current clinical guidelines are in broad agreement that men with a serum level of TT > 12 nmol/l (350 ng/dl) are unlikely to have clinical features of low testosterone and do not generally require treatment.³⁰ This criterion was adopted in the present review.

Studies with eligibility criteria specifying free testosterone or bioavailable testosterone thresholds were eligible for inclusion in the review if they reported mean baseline TT of ≤ 12 nmol/l (or equivalent).

The original protocol for this review further specified the following criteria for trial eligibility, owing to the variability in testosterone assays:

- Clinical symptoms and/or signs of low testosterone (e.g. sexual dysfunction).
- The following information regarding testosterone samples must be available:
 - when samples were collected and assayed (since dates may differ)
 - details of any extraction method used prior to testosterone assay
 - details of the assay method and manufacturer
 - details of any local correction made to adjust assay measurements
 - relevant validation data for the assay (e.g. external quality assurance).

Initial screening of search results suggested that most studies did not report this information and, thus, would subsequently be excluded from the review. In consultation with the Advisory Group for this project, it was decided that such studies were relevant to address the research question of this evidence synthesis and had to be included, if they also fulfilled the remaining eligibility criteria. Therefore, those criteria were subsequently removed to widen inclusion of all relevant studies. Studies assessing any intervention(s) over and above testosterone and placebo were eligible for inclusion, but only data for the testosterone and placebo groups were used in the statistical analyses.

The following eligibility/inclusion criteria were further specified:

- Participants with hypogonadism caused by congenital disorders (e.g. Klinefelter syndrome) or acquired gonadal injury and participants with secondary hypogonadism (hypogonadotropic hypogonadism) were not deemed suitable for inclusion.
- Participants with concomitant medical conditions were considered suitable for inclusion only if their concomitant conditions were those that represent the constellation of low testosterone (i.e. obesity; type 2 diabetes; metabolic syndrome; osteopenia/osteoporosis; and/or history of fracture – for example, bone mineral density *T*-score < 2 and frailty).

Intervention

The intervention considered was TRT [also referred to as androgen replacement therapy (TRT)] with any testosterone formulation, dose, frequency and route of administration (e.g. intramuscular, subdermal, transdermal, oral and buccal preparations of testosterone). Studies involving previous treatment with finasteride or previously unsuccessful treatment with tadalafil were eligible for inclusion. Studies in which participants received other androgens apart from testosterone and studies that allowed concurrent treatment with other hormones or concomitant interventions alongside testosterone were not deemed suitable for inclusion.

Comparator

The eligible comparator treatment was placebo, which was required to be equivalent inactive treatment.

Outcomes

Outcomes of interest included: sexual function, physical health parameters, functional activities, psychological symptoms, CV and cerebrovascular (CBV) events, other comorbidities, prostate-related outcomes, physiological markers, QoL and mortality. The primary and secondary outcomes were identified and agreed upon by the members of the Advisory Group for this project before any statistical analysis was performed.

Primary outcomes

- All-cause mortality.
- Any type of CV and/or CBV event (including fatal events).

Secondary outcomes

- QoL:
 - Short form-36 items/short form-12 items (SF-36/SF-12).
 - Ageing males' symptoms (AMS).
 - WHOLQOL-OLD.
 - Herschbach questionnaire.
- Sexual function:
 - The International Index of Erectile Function – 15 items (IIEF-15).
 - The International Index of Erectile Function – 5 items (IIEF-5).
 - Androgen deficiency in ageing males (ADAM).
 - Psychosexual daily questionnaire (PDQ).

- Derogatis Interview for Sexual Functioning in men-II5 (DISF-II).
- Eleven questions about sexual functioning (ESF).
- Hypogonadism energy diary (HED).
- Sexual Arousal, Interest, and Drive Scale (SAID).
- Male Sexual Health Questionnaire-Ejaculatory Dysfunction-Short Form (MSHQ).
- Physiological markers:
 - Testosterone (nmol/l).
 - Free testosterone (pmol/l).
 - Fasting glucose (mmol/l).
 - Cholesterol (mmol/l).
 - Low-density lipoprotein (LDL; mmol/l).
 - High-density lipoprotein (HDL; mmol/l).
 - Triglycerides (mmol/l).
 - Haemoglobin (Hb; g/l).
 - Glycated haemoglobin (HbA1c; mmol/mol).
 - Haematocrit (%).
 - Systolic blood pressure (SBP; mmHg).
 - Diastolic blood pressure (DBP; mmHg).
 - Areal bone mineral density (BMD).
 - Volumetric bone mineral density.
- Psychological symptoms:
 - Beck depression inventory (BDI).
 - Positive and Negative Affect Scale (PANAS).
 - Hospital Anxiety and Depression Scale (Depression only) HADS-Depression.
 - Patient health questionnaire-9 (PHQ-9).
 - Centre for Epidemiologic Studies Depression Scale (CES-D).
 - Aggression questionnaire.
 - Spielberger State-Trait Anxiety.
 - The Geriatric Depression Scale (GDS).
 - Hamilton Depression and Melancholia Scale.
 - Profile of Mood States (POMS).
- Additional outcomes:
 - Diabetes/diabetes complications.
 - Prostate cancer.
 - Oedema.
 - Hypertension.
 - High haematocrit.
 - Venous thromboembolism.
 - Non-stroke CBV pathology (e.g. carotid occlusion and carotid stenosis).

When multiple assessments were reported, outcomes assessed at 12 months or at the closest time point to 12 months were selected for analysis.

Identifying studies: information sources and search strategy

Highly sensitive search strategies were designed by an information scientist using appropriate subject headings and text word terms to identify reports of published, ongoing and unpublished RCTs reporting the clinical effectiveness of TRT in men with low testosterone. The searches were restricted to reports published from 1992 (year of the first published randomised placebo-controlled study of testosterone administration) to reflect the introduction of TRT in clinical practice and to reports published in English. The Cochrane Highly Sensitive Search Strategy for identifying RCTs was used in MEDLINE and adapted

for other electronic databases. The searches were conducted in August 2018. The following databases were searched to identify relevant clinical trials: MEDLINE, MEDLINE In-process & Other Non-indexed Citations, MEDLINE Epub Ahead of Print, EMBASE, Science Citation Index and the Cochrane Controlled Trials Register (CENTRAL). Cochrane Database of Systematic Reviews (CDSR), Database of Abstracts of Review of Effects (DARE) and the HTA databases were searched for evidence syntheses. Recent conference proceedings of key professional organisations in the fields of endocrinology (e.g. American Endocrine Society), cardiology (e.g. American College of Cardiology) and men's health (e.g. European Menopause and Andropause Society, International Society of Men's Health) were also searched.

Reference lists of all included studies were perused to identify additional potentially relevant reports. We also contacted our panel of experts for details of any additional potentially relevant reports. A full MEDLINE search strategy is presented in [Appendix 1](#).

Ongoing studies were identified through searching Current Controlled Trials, Clinical Trials and World Health Organization (WHO) International Clinical Trials Registry. Websites of professional organisations, regulatory bodies and health technology assessment (HTA) organisations were also searched to identify additional relevant reports.

Study selection process

Two reviewers independently screened titles and abstracts of all citations identified by the search strategies (MC and MB or MA-M). All potentially relevant reports were retrieved in full and assessed by one reviewer (MC) with 10% independently checked by a second reviewer (MA-M). In addition, all selected reports were independently assessed by a clinical expert (CJ or RQ). Any disagreements during the selection process were resolved by consensus.

Data collection processes

Aggregate data

A data extraction form was designed specifically for this assessment to collect aggregated data from all included studies, regardless of whether IPD could be obtained. Following piloting and further refinement of the form, one reviewer (MC) extracted details of study design, characteristics of studies, interventions and participants and outcome measures. A second reviewer (MA-M) cross-checked a random sample of 10% of selected studies. Extracted data were further checked for accuracy by the project statistician (JH). Any disagreements were resolved by consensus. Details of the items collected in the data extraction form are presented in [Appendix 2](#).

Individual participant data

Anonymised data for each of the pre-specified variables were required for each randomised participant from as many identified studies as possible.

We established a collaborative group of the investigators of all identified trials. Contact information of authors of each eligible study was identified from the published report(s) and by electronic searches. All first authors of all eligible studies were initially contacted by email with a brief summary of the scope, rationale and objectives of the project, and invited to join the collaboration and share their anonymised IPD. Reminders were sent to non-responders after 1 week. Where reminders did not elicit a response, other methods of communication were attempted, including telephone calls, letter by post and attempts to contact other investigators of the respective studies. Once a memorandum of understanding was obtained, preferably electronically, investigators were asked to agree to share and transfer their trial data by signing a Data-Sharing Agreement form, which specified that only anonymised data were requested and accepted, stored securely and used exclusively for the purposes of the project. Trial data were considered unavailable in the event that no study authors had responded to multiple contact attempts or when the trial authors indicated that they had no longer access to the trial data. Trial authors were

requested to provide data on demographic and baseline characteristics of participants and relevant outcomes (see 'List of core items' in [Appendix 3](#) for further details).

A standard operating procedure (SOP) was developed specifically for the management of the IPD (see [Appendix 4](#)). A designated Gateway Manager (MC) was responsible for receiving the anonymised IPD from study collaborators and ensuring secure storage on a password-protected computer server area. Access to data at all stages of data process (checking and cleaning) and analysis was restricted to core members of the research team (MC, JH, MB, HR). Transfer of data to personal devices was forbidden at any time. Data were received in a variety of formats according to the security requirements of the countries from which data were received. In most cases, data transfer took place via an electronic system in which files were securely encrypted.

Data items and individual participant data integrity

Data requested for each randomised participant are presented in [Appendix 3](#).

Data sets received from the collaborators were initially checked to ensure they were from the correct identified trials and in satisfactory condition to be included in the analyses. Data cleaning was carried out at the individual data set level prior to being merged into a master database. Within each data set, variable names were standardised and checked for accuracy using summary statistics with published data. Clarification was sought from the trial's authors when discrepancies arose. If clarification was not available or successful, the research team discussed any major discrepancies and decided whether data were eligible for inclusion. After data cleaning was complete, the physiological markers were standardised to be on the same scale. CV and CBV variables, and the additional secondary outcomes, were categorised independently by two clinical review authors (CJ, RQ). Any disagreements were resolved by discussion. CV events were based on standardised definition.³² The IPD master database comprising all data sets obtained from the collaborators was created using the Stata statistical software version 16.³³

Risk of bias assessment in individual studies

Two reviewers (MC, MA-M) independently assessed risk of bias of all included studies using the original version of the Cochrane Collaboration's risk of bias tool for randomised trials.³⁴ For the studies that provided IPD, follow-up enquiries were made with the respective collaborator where details required to assess any domain(s) were unclear or not reported, and the risk of bias assessment was updated accordingly with any additional information. We also conducted risk of bias assessment based on information reported in the trial publications. No follow-up enquiries were made regarding missing or unclear information for studies that did not provide IPD. The following domains were assessed: random sequence generation, allocation concealment, blinding of participants, blinding of personnel, blinding of outcome assessment, incomplete outcome data, selective outcome reporting and any other bias (e.g. conflict of interest, contamination bias). Individual items were categorised as high risk of bias, low risk of bias or unclear risk of bias. In particular, blinding of participants/personnel/outcome assessors was judged to be HIGH if any member of the study team was not blinded. Studies reported as 'double blinded' were assessed as UNCLEAR if no further details of blinding were provided.

Selective outcome reporting was assessed as UNCLEAR if the protocol was not available for cross-checking; other bias was judged to be HIGH if the study was sponsored and/or conducted by a pharmaceutical company. We considered a study to have an overall (1) HIGH risk of bias if one or more key domains (selection bias, detection bias) were judged to be at high risk; (2) UNCLEAR risk of bias if one or more key domains were judged to be at unclear risk; (3) LOW risk of bias if all key domains were judged to be at low risk. Any disagreements were resolved by consensus between review authors.

Statistical analyses

All analyses were conducted according to the intention-to-treat (ITT) principle, following a pre-specified statistical analysis plan and undertaken using Stata 16.³³ Treatment effects are presented with 95%

confidence intervals (CIs) unless otherwise stated. *P*-values for the primary outcomes are reported in tables and in the text, while those for the secondary outcomes are only reported in the text. No adjustment for multiple secondary outcomes was performed. As participants could experience more than one event, the analyses for primary outcomes were conducted at the participant level and not according to the number of events.

Baseline characteristics and outcome measures are described by randomisation group using appropriate summary statistics.

For the main analyses, we used two approaches: a one-stage meta-analysis approach for each outcome using the acquired IPD and a two-stage approach enabling the integration of the IPD along with the extracted published study summaries for studies eligible but without IPD.

To allow direct comparison, SF-36 and SF-12 scores were transformed into *T*-scores.³⁵

For the primary outcomes, to assess small-study effects and publication biases we used counter-enhanced funnel plots and tests for asymmetry using Peter's test.

One-stage analyses

A one-stage meta-analysis approach involves fitting a regression model to the entire IPD data set rather than to each trial data set separately. This model accounts for the clustering of participants within trials. For the primary outcomes (all binary), a fixed-effect logistic regression model accounting for clustering and allowing a separate intercept per study was used. This is because a random-effects model – with a random intercept on study and random slope on treatment – failed to converge. Effect estimates were presented as odds ratios (ORs) and accompanying 95% CIs. For the secondary outcomes (all continuous), a random-effects linear regression model with a random intercept on study and a random slope on treatment was performed, accounting for clustering and allowing separate baseline adjustment per study as well as a separate residual variance using restricted maximum likelihood (REML). These effect estimates were presented as mean differences (MDs) and accompanying 95% CIs. The estimated between-study variance, τ^2 , is reported to assess heterogeneity. The additional outcomes are summarised descriptively with a post hoc chi-squared analysis undertaken.

Two-stage analyses

Given that IPD were not available for all eligible studies, two-stage meta-analyses were also undertaken on all viable outcomes. Outcomes were analysed in their original trial and then combined in a meta-analysis to give an overall measure of effect. The first stage involves analysing IPD separately in each study to obtain aggregate data (i.e. the treatment effect in each study). For the primary outcomes, logistic regression models were fitted for each outcome. For secondary outcomes, linear regression models were similarly fitted as well as adjusting for baseline score. For studies with no IPD, we obtained effect estimates and standard errors according to current methodological recommendations, either directly from study publications or through communication with studies' authors, as aggregated estimates.³⁴ In the second stage, the effect estimates from the IPD and aggregate studies were pooled together using a random-effects model with REML to produce effect estimates by study, IPD studies, aggregate data and overall. For models that would not converge using REML, then a random effects model using the DerSimonian and Laird method was used. Heterogeneity was assessed using the I^2 statistic.

Sensitivity analyses

Due to the low event rate for the primary outcome, mortality, a Mantel–Haenszel method was also performed for the two-stage analysis.^{34,36} A sensitivity analysis was also performed for the primary outcome, CV and/or CBV events, by including unknown cause of death using the same analysis as described under the one-stage analysis.

For the secondary outcomes, glucose and HbA1c, a sensitivity analysis of removing participants with diabetes at baseline was performed as these participants could have been on medication which artificially lowers these outcomes.

Pre-specified subgroup analyses

Pre-specified subgroup analysis was conducted to explore possible treatment-modifying effects of diabetes, smoking status, testosterone and free testosterone levels on the primary outcome, CV and/or CBV events. A subgroup analysis for mortality proved unfeasible due to the limited number of reported events across trials. We originally planned to specify categories for testosterone and free testosterone; however, due to the sparse data available within each category, these were explored as continuous variables.

Subgroup by treatment interactions were assessed by including the within-interaction terms in the models outlined above. As for current methodological recommendations, continuous covariates were centred on the mean value within each trial and binary covariates were centred on the proportion within each trial.³⁷ Subgroup analyses were performed on the IPD studies.

A stricter level of statistical significance (two-sided 1% significance level) was applied given their exploratory nature and corresponding CIs were set at 99%.

Post hoc subgroup analyses

A post hoc analysis on possible modifying effects of age on CV and/or CBV events was undertaken. Post hoc analyses on the secondary outcomes were also undertaken following the same method described for the pre-specified subgroup analyses. It was decided to perform analyses on the most reported outcomes by the trials included in the IPD analysis or the outcomes with substantial heterogeneity (τ^2) compared to other outcomes. The selected outcomes were AMS for QoL, IIEF-15 and IIEF-5 for sexual function, Hb for physiological markers and BDI for psychological symptoms.

We also performed a post hoc subgroup analysis based on IIEF-15 and its subscales assessing the effects of age, total serum testosterone and body mass index (BMI).

Threshold analysis

A regression analysis was performed to see whether there were any thresholds for IIEF-15 at follow-up for age, baseline total serum testosterone and BMI and for IIEF-15 at baseline for total serum testosterone. According to the number of categories identified, we performed either an analysis of variance (ANOVA) or *t*-test statistical analysis to confirm whether the categories were significant.

Results of the quantitative synthesis and individual participant data meta-analysis

Study selection and individual participant data obtained

The literature searches identified 9871 records. After de-duplication, 5603 records were screened for relevance. Of these, 225 were considered potentially relevant and selected for full-text assessment. Of the 225 articles retrieved and assessed in depth, 109 publications reporting 35 studies met the inclusion criteria, while the remaining 116 articles were deemed not suitable for inclusion. We sought IPD from all 35 studies (total of 5601 randomised participants) and obtained IPD from 17 studies (total of 3431 out of 3474 randomised participants in these 17 studies). The IPD integrity section reports details of discrepancies between numbers of randomised participants reported in publications and IPD received. Aggregate data were available for all 35 studies (5601 randomised participants). The PRISMA-IPD flow diagram illustrating the study selection process is presented in [Figure 1](#).

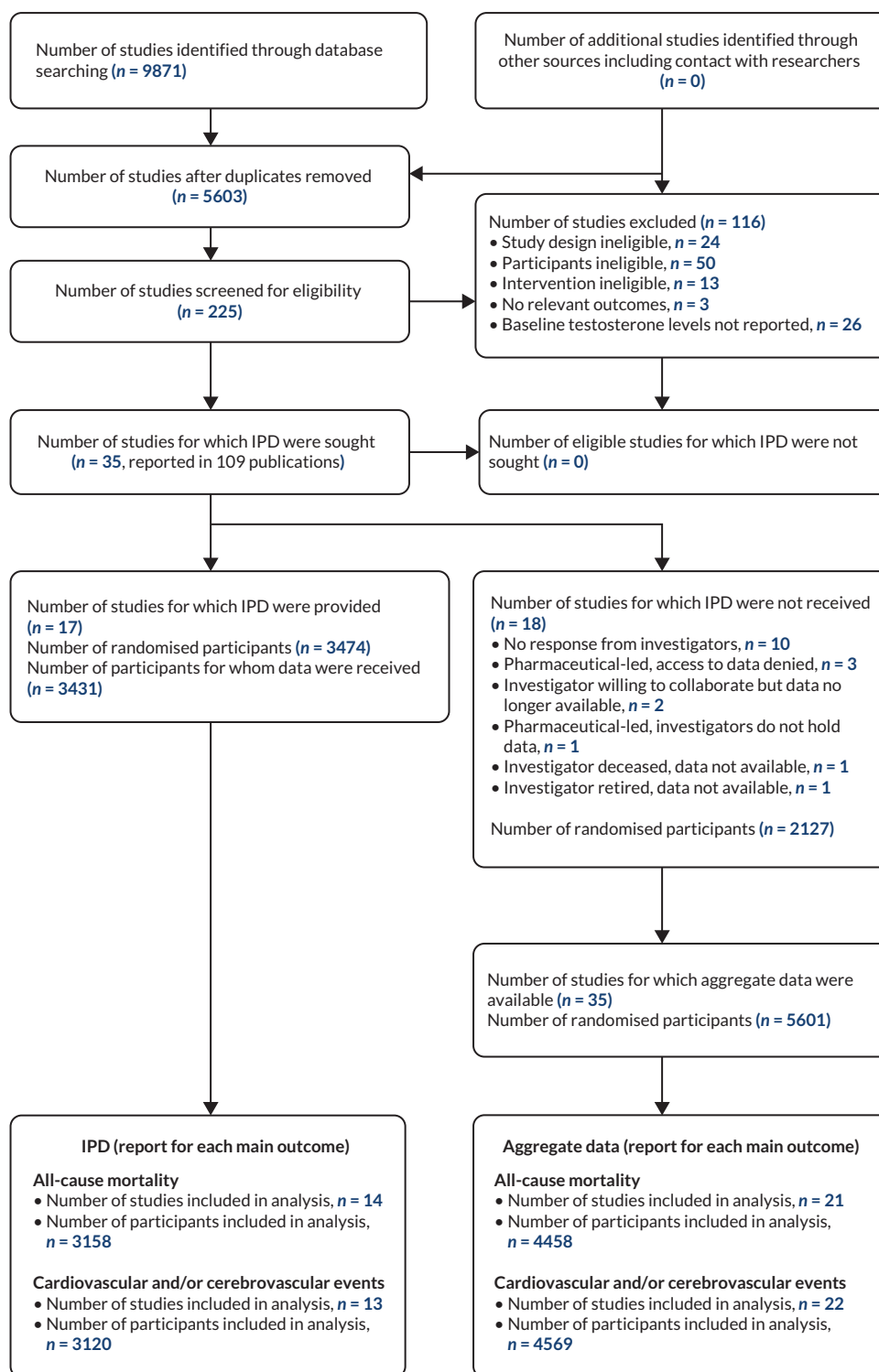


FIGURE 1 PRISMA-IPD flow diagram of selected studies.

Study characteristics

The study characteristics of the 35 included RCTs (total number of participants randomised to testosterone or placebo, as reported in the respective publications: 5601) are detailed in [Appendix 5, Table 31](#) (studies providing IPD) and [Appendix 5, Table 32](#) (studies not providing IPD) and summarised in [Table 1](#) (studies providing IPD) and [Table 2](#) (studies not providing IPD). These tables include only information from the relevant publications, and not from any IPD received from collaborators.

TABLE 1 Summary of the characteristics of the 17 studies that provided IPD

Study ID	Geographical location	No. of centres	Total <i>n</i> randomised	Treatment duration/ length of follow-up	Testosterone assay as reported by the study authors
Amory 2004 ³⁸	USA	1	48	3 years/ 3 years ^a	Fluoroimmunoassay (Delfia, Wallac Oy, Turku, Finland)
Basaria 2010 ¹²	USA	3	209	6 months/ 6 months	Immunoassay (Quest)
Basaria 2015 ³⁹	USA	3	308	3 years/ 3 years ^b	Bayer Advia Centaur immunoassay (Siemens Healthcare Diagnostics)
Brock 2016 ⁵²	Argentina, Canada, Germany, Spain, Italy, South Korea, Puerto Rico, UK, USA	98	715	12 weeks/ 12 weeks	Liquid chromatography-mass spectrometry/mass spectrometry
Emmelot-Vonk 2008 ⁴⁵	Netherlands	1	237	6 months/ 6 months	Solid-phase, competitive, chemiluminescent enzyme immunoassay (Immulite 2000, Diagnostic Products Corporation, Los Angeles, CA, USA)
Gianatti 2014 ⁴⁶	Australia	1	88	30 weeks/ 40 weeks	ECLIA and LCMS/MS
Giltay 2010 ⁴⁷	Russia	1	184	30 weeks/ 30 weeks	Vitros 3600 system (Ortho-Clinical Diagnostics, Johnson & Johnson company, New Brunswick, NJ, USA) with a chemiluminescence immunoassay technology
Groti 2018 ⁴⁸	Slovenia	1	55	12 months/ 12 months	Coated tube RIA (DiaSorin S. p. A., Salluggia, Italy and Diagnostic Products Corporation, Los Angeles, CA, USA)
Hackett 2013 ⁴²	UK	8	199	18 weeks/ 30 weeks	Roche common platform immunoassay
Hildreth 2013 ⁴⁰	USA	1	83	12 months/ 12 months	ELISA using a Beckman Coulter (Brea, CA, USA) Access II analyser
Ho 2012 ⁴⁹	Malaysia	1	120	42 weeks/ 48 weeks	Immunoassay using an AxSYM testosterone assay (Abbott Laboratories, Wiesbaden, Germany), based on microparticle enzyme immunoassay technology
Magnussen 2016 ⁵⁰	Denmark	1	43	24 weeks/ 24 weeks	Liquid chromatography tandem mass spectrometry after ether extraction (Statens Serum Institut, Copenhagen, Denmark)
Marks 2006 ⁴¹	USA	1	44	6 months/ 6 months	Mass spectroscopy
Merza 2006 ⁴³	UK	1	39	6 months/ 6 months ^c	IRMA (Orion Diagnostics)

continued

TABLE 1 Summary of the characteristics of the 17 studies that provided IPD (*continued*)

Study ID	Geographical location	No. of centres	Total <i>n</i> randomised	Treatment duration/length of follow-up	Testosterone assay as reported by the study authors
Snyder 2016 ¹¹	USA	12	790	12 months/ 12 months	Liquid chromatography with tandem mass spectroscopy
Srinivas-Shankar 2010 ⁴⁴	UK	1	274	6 months/ 6 months	Chemiluminescent immunoassay with a Roche Elecys E170 platform
Svartberg 2008 ⁵¹	Norway	1	38	40 weeks/ 52 weeks	Electrochemical luminescence immunoassay using an automated clinical chemistry analyser (Modular E; Roche Diagnostics GmbH, Mannheim, Germany)

ELISA, enzyme-linked immunosorbent assay, IRMA, immunoradiometric assay, RIA, radioimmunoassay.

a Outcome data were used at 12 months.

b Secondary outcome data were used at 18 months and primary outcome data at 3 years due to being unable to confirm the time frame in which these events occurred.

c Plus a further 6-month open-label phase, not included in this review.

Note

The manufacturer of testosterone assay method was not reported for all studies.

TABLE 2 Summary of the characteristics of the 18 studies that did not provide IPD

Study ID	Geographical location	No. of centres	Total <i>n</i> randomised	Treatment duration/length of follow-up	Testosterone assay as reported by the study authors
Aversa 2010a ⁶⁰	Italy	1	50	12 months/ 12 months ^a	Electrochemiluminescence (Immolute 2000 Siemens, Milan, Italy)
Aversa 2010b ⁶¹	Italy	NR	52	12 months/ 12 months	Electrochemiluminescence (Immolute 2000 Siemens, Milan, Italy)
Basurto 2008 ⁶³	Mexico	1	48	12 months/ 12 months	Specific solid-phase radioimmunoassay (Diagnostic Products Corporation, Los Angeles, CA, USA)
Behre 2012 ⁶⁸	Austria, Finland, Germany, Ireland, Italy, Spain, Sweden, UK	NR	362	6 months/ 6 months ^b	Electrochemiluminescence immunoassay technique on a Roche Elecys or Modular E170 analyser
Borst 2014 ⁵³	USA	NR	30	12 months/ 12 months	Electrochemiluminescence immunoassay (Cobas)
Cavallini 2004 ⁶²	Italy	NR	85	6 months/ 6 months	Recombinant immunoassay after extraction and celite chromatography (Diagnostic Products, Los Angeles, CA, USA)
Cherrier 2015 ⁵⁴	USA	1	22	6 months/ 6 months	Liquid chromatography tandem mass spectrometry
Chiang 2007 ⁶⁴	Taiwan	2	40	3 months/ 3 months	Radioimmunoassay
Clague 1999 ⁶⁵	UK	1	14	12 weeks/ 12 weeks	NR

TABLE 2 Summary of the characteristics of the 18 studies that did not provide IPD (*continued*)

Study ID	Geographical location	No. of centres	Total <i>n</i> randomised	Treatment duration/length of follow-up	Testosterone assay as reported by the study authors
Dhindsa 2016 ⁵⁵	USA	1	44	22 weeks/ 24 weeks	Liquid chromatography tandem mass spectrometry (Quest Diagnostics)
Dias 2016 ⁵⁶	USA	1	29	12 months/ 12 months	Liquid chromatography tandem mass spectroscopy
Jones 2011 ⁶⁹	Belgium, France, Germany, Italy, Netherlands, Spain, Sweden, UK	36	220	12 months/ 12 months	NR
Kaufman 2011 ⁵⁷	USA	63	274	182 days/ 182 days	Liquid chromatography tandem mass spectrometry (Pharmaceutical Product Development, Richmond, VA, USA)
Kenny 2010 ⁵⁸	USA	1	131	12 months/ 12 months ^c	Radioimmunoassay (Endocrine Sciences Inc, Calabasas Hills, CA, USA)
Morales 2009 ⁶⁶	Canada	4	58	4 months/ 4 months	NR
Paduch 2015 ⁷⁰	USA, Canada, Mexico	NR	76	16 weeks/ 16 weeks	Liquid chromatography tandem mass spectrometry
Steidle 2003 ⁵⁹	USA	43	406	90 days/ 90 days	Radioimmunoassay (Diagnostic Products, Los Angeles, CA, USA)
Wang 2013 ⁶⁷	China	1	186	24 months/ 24 months	Chemical luminescence

NR, not reported.

a Plus 12 months open-label treatment, not included in this review.

b Plus 12 months open-label extension, not included in this review.

c The study was designed as a 2-year intervention but limited 24-month follow-up was available due to recruitment delays and funding issues.

Note

Manufacturer of testosterone assay method was not reported for all studies.

Individual participant data studies

Of the 17 studies that provided data for the IPD analyses, 6 were conducted in the USA,^{11,12,38–41} 3 in the UK,^{42–44} 1 in each of the Netherlands,⁴⁵ Australia,⁴⁶ Russia,⁴⁷ Slovenia,⁴⁸ Malaysia,⁴⁹ Denmark⁵⁰ and Norway,⁵¹ and the remaining study was conducted across nine countries (Argentina, Canada, Germany, Spain, Italy, South Korea, Puerto Rico, UK, USA).⁵² The majority of studies were single centre,^{38,40,41,43–51} two studies involved 3 centres,^{12,39} one study involved 8 centres,⁴² one study 12 centres,¹¹ and another study 98 centres.⁵² Across studies, the total numbers of randomised participants ranged from 27⁵¹ to 790¹¹ and duration of treatment from 12 weeks⁵² to 3 years.^{38,39}

Non-IPD studies

Seven of the 18 studies that did not provide data for the IPD analyses were conducted in the USA;^{53–59} 3 studies were conducted in Italy;^{60–62} 1 study was conducted in each of Mexico,⁶³ Taiwan,⁶⁴ UK,⁶⁵ Canada⁶⁶ and China;⁶⁷ and 3 studies were conducted in multiple countries: Austria, Finland, Germany, Ireland, Italy, Spain, Sweden, UK;⁶⁸ Belgium, France, Germany, Italy, Netherlands, Spain, Sweden, UK,⁶⁹ and USA, Canada, Mexico.⁷⁰ Eight of the 18 studies were single-centre studies.^{54–56,60,63,65} One study

was conducted in 2 centres (40 randomised participants),⁶⁴ and one study in each of 4 centres (58 randomised participants),⁶⁶ 36 centres (220 randomised participants),⁶⁹ 43 centres (406 randomised participants)⁵⁹ and 63 centres (274 randomised participants).⁵⁷ The remaining five studies did not report the number of centres involved. Total numbers of randomised participants ranged from 14⁶⁵ to 406⁵⁹ and treatment duration from 3 months^{57,59,64,65} to 24 months.⁶⁷ Three studies that did not provide data for the IPD analysis were linked to pharmaceutical companies;^{59,68,69} two of these studies provided full disclosure of all serious adverse events (SAEs).^{68,69}

Participant characteristics

The participant characteristics of the 35 included RCTs are detailed in [Appendix 5, Table 33](#) (studies providing IPD) and [Appendix 5, Table 34](#) (studies not providing IPD) and summarised in [Table 3](#) (studies providing IPD) and [Table 4](#) (studies not providing IPD).

Individual participant data studies

Mean age of the testosterone and placebo groups was reported in 14 of the 17 studies that provided IPD^{11,12,38-40,42-45,47,49-52} and ranged from 51.6 years⁴⁷ to 74 years¹² years in the TRT group and from 52.8 years⁴⁷ to 74 years¹² years in the placebo group. One study reported an overall mean age of 60.2 years for all participants.⁴⁸ One study reported a median age of 62 years in both treatment groups⁴⁶ and another study reported a median age of 68 years in the TRT group and 70 years in the placebo group.⁴¹

Mean BMI was reported in 13 studies^{11,12,38,39,42,44,45,47-52} and ranged from 27.4⁴⁵ to 35.3⁴⁷ in the active treatment groups and from 27.3⁴⁵ to 34.2⁴⁷ in the placebo groups. Of these 13 studies, mean BMI was in the obese range (i.e. ≥ 30) in at least one arm of nine studies.^{11,12,42,47-52} Median BMI was reported by two studies and was 31.5 and 33.4 in the active and placebo groups, respectively, in one study⁴⁶ and 28.3 and 29.6, respectively, in the other study.⁴¹ The remaining two studies did not report information on BMI.^{40,43}

Nine studies reported mean baseline testosterone in terms of nmol/l^{38,42-45,47-49,51} and five studies in units of ng/dl,^{11,12,39,40,52} which were converted to nmol/l for consistency.⁷¹ Mean baseline testosterone in the TRT groups ranged from 6.7⁴⁷ to 11 nmol/l^{44,45} and the placebo groups ranged from 7⁵² to 10.9 nmol/l.⁴⁴ Three studies reported median testosterone, ranging from 7.1⁵⁰ to 8.7 nmol/l⁴⁶ in the TRT groups and from 8.5⁴⁶ to 9.4 nmol/l⁵⁰ in the placebo groups.

Overall, the baseline characteristics of participants included in the IPD were well balanced between TRT and the placebo groups (see [Table 5](#)). The mean age was 65 years [standard deviation (SD) 11 years, 16 studies] with a mean BMI of 30 kg/m² (SD 5 kg/m², 17 studies). The majority of the participants were white (88%, six studies) and non-smokers (TRT 89%, placebo 87%, 10 studies). None of the participants had a diagnosis of prostate cancer. Further baseline characteristics can be seen in [Appendix 6, Table 35](#).

Non-individual participant data studies

Mean age was reported in 17 of the 18 studies that did not provide IPD.^{20,55-70} In the active treatment groups, mean age ranged from 47.9 years⁶⁴ to 77.9 years,⁵⁸ and in the placebo groups from 52.7 years⁷⁰ to 54.5 years.⁵⁵ The remaining study did not provide information on the age of participants.⁵⁴

Mean BMI was reported by 13 studies.^{53,55-61,63,66-68,70} In the active treatment groups, mean BMI ranged from 27.2⁵⁸ to 39⁵⁵ and in the placebo groups from 26.6⁵⁸ to 39.4.⁵⁵ Of these 13 studies, mean BMI was in the obese range (i.e. ≥ 30) in at least one arm of eight studies.^{55-57,59-61,66,70} Five studies did not report baseline BMI values.^{54,62,64,65,69}

TABLE 3 Summary of the characteristics of participants enrolled in the 17 studies that provided IPD

Study ID	Study group	N reported in baseline characteristics	Age, years, mean (SD)	BMI, kg/m ² , mean (SD)	TT, mean, nmol/l (SD where reported in nmol/l)	Type 2 diabetes, n (%)
Amory 2004 ³⁸	TRT IM	24	71 (4)	28.7 (3.6)	9.9 (1.6)	NR
	Placebo	24	71 (5)	27.9 (3.6)	10.5 (1.7)	NR
Basaria 2010 ¹²	TRT gel	106	74 (6)	29.7 (4.1)	8.7	NR
	Placebo	103	74 (4)	30.0 (4.2)	8.2	NR
Basaria 2015 ³⁹	TRT gel	155	66.9 (5.0)	28.1 (2.1)	10.6	22 (14.2) (type NR)
	Placebo	151	68.3 (5.3)	28.0 (2.9)	10.6	24 (15.9) (type NR)
Brock 2016 ⁵²	TRT solution	358	54.7 (10.6)	30.3 (4.1)	7	NR
	Placebo	357	55.9 (11.4)	30.9 (4.2)	7	NR
Emmelot-Vonk 2008 ⁴⁵	TRT capsules	113	67.1 (5.0)	27.4 (3.8)	11.0 (1.9)	NR
	Placebo	110	67.4 (4.9)	27.3 (3.9)	10.5 (1.9)	NR
Gianatti 2014 ⁴⁶	TRT IM	45	Median (IQR) 62 (58–68)	Median (IQR) 31.5 (28.3–35.5)	Median (IQR) 8.7 (7.1–11.1)	45 (100)
	Placebo	43	Median (IQR) 62 (57–67)	Median (IQR) 33.4 (31.4–35.4)	Median (IQR) 8.5 (7.2–11.0)	43 (100)
Giltay 2010 ⁴⁷	TRT IM	113	Mean (95% CI) 51.6 (49.8 to 53.4)	Mean (95% CI) 35.3 (34.2 to 36.6)	Mean (95% CI) 6.7 (6.0 to 7.4)	32 (28.3)
	Placebo	71	Mean (95% CI) 52.8 (50.5 to 55.0)	Mean (95% CI) 34.2 (32.9 to 35.7)	Mean (95% CI) 7.5 (6.6 to 8.5)	24 (33.8)
Groth 2018 ⁴⁸	TRT IM	28	Overall mean (SD) 60.2 (7.2)	34.0 (4.4)	7.2 (2.0)	NR
	Placebo	27		32.6 (3.7)	8.0 (1.3)	NR
Hackett 2013 ⁴²	TRT IM	92	61.2 (10.5)	33.0 (6.1)	9.2 (3.5)	92 (100)
	Placebo	98	62.0 (9.3)	32.4 (5.5)	8.9 (3.5)	98 (100)
Hildreth 2013 ⁴⁰	TRT gel	55	66.4 (5.0)	NR	10.4	NR
	Placebo	28	67.5 (5.6)	NR	10.4	NR
Ho 2012 ⁴⁹	TRT IM	60	53.4 (7.4)	30.4 (5.2)	8.9 (2.0)	14 (23.3) (type NR)
	Placebo	60	53.0 (8.2)	28.2 (4.5)	9.1 (1.8)	9 (15) (type NR)

continued

TABLE 3 Summary of the characteristics of participants enrolled in the 17 studies that provided IPD (continued)

Study ID	Study group	N reported in baseline characteristics	Age, years, mean (SD)	BMI, kg/m ² , mean (SD)	TT, mean, nmol/l (SD where reported in nmol/l)	Type 2 diabetes, n (%)
Magnussen 2016 ⁵⁰	TRT gel	20	61 (6)	Arithmetic mean (IQR) 30.6 (28.9–32.2)	Median (IQR) 7.1 (6.6–11.9)	20 (100)
	Placebo	19	59 (6)	Arithmetic mean (IQR) 30.8 (28.9–32.6)	Median (IQR) 9.4 (8.1–12.5)	19 (100)
Marks 2006 ⁴¹	TRT IM	21	Median (range) 68 (44–78)	Median (range) 28.3 (22.7–37.9)	Median (range) 7.7 (5.7–11.1)	NR
	Placebo	19	Median (range) 70 (45–78)	Median (range) 29.6 (23.6–37.8)	Median (range) 8.7 (5–11.4)	NR
Merza 2006 ⁴³	TRT patch	20	63 (9)	NR	8.4 (3.3)	NR
	Placebo	18	59.7 (10.2)	NR	7.5 (2.5)	NR
Snyder 2016 ¹¹	TRT gel	394	72.1 (5.7)	31.0 (3.5)	8	148 (37.5) (type NR)
	Placebo	394	72.3 (5.8)	31.0 (3.6)	8.2	144 (36.5) (type NR)
Srinivas-Shankar 2010 ⁴⁴	TRT gel	130	73.7 (5.7)	27.9 (4.1)	11.0 (3.2)	NR
	Placebo	132	73.9 (6.4)	27.7 (4.0)	10.9 (3.1)	NR
Svartberg 2008 ⁵¹	TRT IM	17	69 (5)	30.6 (3.8)	8.4 (1.7)	NR
	Placebo	18	69 (5)	28.6 (3.7)	8.2 (2.1)	NR

IM, intramuscular; NR, not reported.

Note

 Testosterone was converted to nmol/l for studies reporting ng/dl.⁷¹ SD is not shown for these studies.

TABLE 4 Summary of the characteristics of participants enrolled in the 18 studies that did not provide IPD

Study ID	Study group	N reported in baseline characteristics	Age, years, mean (SD)	BMI, kg/m ² , mean (SD)	TT, mean, nmol/l (SD where reported in nmol/l)	Type 2 diabetes, n (%)
Aversa 2010 ⁶⁰	TRT IM	40	58 (10)	30.2 (4.5)	8.3 (2.4)	NR
	Placebo	10	57 (8)	31 (6.2)	9.0 (1.7)	NR
Aversa 2010 ⁶¹	TRT oral ^a	10	57 (8)	32.5 (5.2)	NR	3 (30.0)
	TRT IM	32	58 (10)	30.2 (4.5)	NR	10 (31.2)
	Placebo	10	55 (5)	31 (6.2)	NR	4 (40.0)
Basurto 2008 ⁶³	TRT IM	25	63.2 (7.9)	27.4 (3.0)	10.4	NR

TABLE 4 Summary of the characteristics of participants enrolled in the 18 studies that did not provide IPD (*continued*)

Study ID	Study group	N reported in baseline characteristics	Age, years, mean (SD)	BMI, kg/m ² , mean (SD)	TT, mean, nmol/l (SD where reported in nmol/l)	Type 2 diabetes, n (%)
Behre 2012 ⁶⁸	Placebo	23	63.1 (7.7)	27.2 (2.0)	10.7	NR
	TRT gel	183	61.9 (6.6)	28.5 (3.3)	10.4 (2.6)	NR
Borst 2014 ⁵³	Placebo	179	62.1 (6.3)	28.7 (3.0)	10.6 (2.6)	NR
	TRT IM	14 ^b	69.2 (8.0)	29.4 (4.6)	10.4 (2.6)	NR
Cavallini 2004 ⁶²	Placebo	16 ^b	70.8 (9.7)	28.7 (3.0)	10.6 (2.6)	NR
	TRT oral	40	64 (range 60–72)	NR	9.9 (1.8)	NR
Cherrier 2015 ⁵⁴	Placebo	45	63 (range 61–74)	NR	10.5 (2.1)	NR
	TRT gel	12	NR	NR	10.7	NR
Chiang 2007 ⁶⁴	Placebo	10	NR	NR	9.8	NR
	TRT gel	20 ^b	47.9 (17.0)	NR	7.4	NR
Clague 1999 ⁶⁵	Placebo	18 ^b	56.1 (14.6)	NR	9.1	NR
	TRT IM	7	68.1 (6.6)	NR	11.3 (1.7)	NR
Dhindsa 2016 ⁵⁵	Placebo	7	65.3 (1.8)	NR	11.6 (0.9)	NR
	TRT IM	20	54.7 (7.8)	39.0 (7.6)	8.7	22 (100)
Dias 2016 ⁵⁶	Placebo	14	54.5 (8.7)	39.4 (7.9)	8.3	22 (100)
	TRT gel	13	72 (SEM 1)	30.1 (SEM 1.1)	10.4	NR
Jones 2011 ⁶⁹	Placebo	9	72 (SEM 1)	27.6 (SEM 1.2)	10.8	NR
	TRT gel	108	59.9 (9.1)	NR	9.2 (2.6)	68 (63.0)
Kaufman 2011 ⁵⁷	Placebo	112	59.9 (9.4)	NR	9.5 (3.3)	69 (61.6)
	TRT gel	214	53.6 (9.5)	31.3 (4.2)	9.8	NR
Kenny 2010 ⁵⁸	Placebo	37	55.5 (10.3)	30.6 (4.1)	10.2	NR
	TRT gel	69	77.9 (7.3)	27.2 (4.3)	13.2	12 (17.4) (type NR)
Morales 2009 ⁶⁶	Placebo	62	76.3 (8.0)	26.6 (4.2)	14.5	10 (16.1) (type NR)
	TRT capsules	24	59 (10.6)	31.3 (5.4)	10.2 (4.9)	NR
Paduch 2015 ⁷⁰	Placebo	28	60.2 (9.6)	29.7 (4.4)	10.0 (5.5)	NR
	TRT solution	40	48.4 (9.8)	30.6 (3.1)	7.4	NR
Steidle 2003 ⁵⁹	Placebo	36	52.7 (9.3)	30.8 (3.2)	7.7	NR
	TRT gel 50 mg	99	58.1 (9.7)	30.0 (3.7)	8.1 (2.0)	NR

continued

TABLE 4 Summary of the characteristics of participants enrolled in the 18 studies that did not provide IPD (*continued*)

Study ID	Study group	N reported in baseline characteristics	Age, years, mean (SD)	BMI, kg/m ² , mean (SD)	TT, mean, nmol/l (SD where reported in nmol/l)	Type 2 diabetes, n (%)
Wang 2013 ⁶⁷	TRT gel 100 mg	106	56.8 (10.6)	29.9 (3.3)	8.1 (2.2)	NR
	TRT patch	102	60.5 (9.7)	29.9 (3.8)	8.3 (2.4)	NR
	Placebo	99	56.8 (10.8)	30.2 (3.8)	7.9 (2.8)	NR
	TRT capsules 40 mg	62	68.1 (5.4)	27.9 (3.2)	7.5	NR
	TRT capsules 20 mg	62	68.4 (5.5)	28.2 (3.6)	7.6	NR
	Placebo	62	68.0 (4.8)	28.7 (2.9)	7.6	NR

IM, intramuscular; NR, not reported; SEM, standard error of mean.

a Group not included in 12 months analysis as study group crossed over at 6 months and therefore was not eligible.

b Number of participants in baseline characteristics not explicitly reported. Testosterone was converted to nmol/l for studies reporting ng/dl.⁷¹ SD is not shown for these studies.

Note

Aversa 2010:⁶¹ mean testosterone across all groups at baseline < 320 ng/dl.

TABLE 5 Baseline characteristics of the participants enrolled in the 17 studies that contributed to the IPD analysis

Baseline characteristics	Number of studies	TRT	Placebo
Age (years)	16	64.5 (11.0); 1724	65.3 (10.8); 1656
BMI (kg/m ²)	17	30.3 (4.7); 1746	30.2 (4.5); 1677
Ethnicity	6		
White		915 (87.5)	888 (87.6)
Asian		63 (6.0)	62 (6.1)
Black/African American		16 (1.5)	12 (1.2)
Other		9 (0.9)	7 (0.7)
Missing		43 (4.1)	45 (4.4)
Smoking status	10		
No		838 (88.9)	756 (87.2)
Yes		103 (10.9)	107 (12.3)
Missing		2 (0.2)	4 (0.5)
Albumin (g/l)	9	42.6 (3.2); 817	42.7 (3.1); 783
Estradiol (pmol/l)	8	80.8 (38.6); 782	77.1 (33.6); 710
Follicle-stimulating hormone (IU/l)	8	14.7 (16.7); 711	14.2 (16.0); 683
Luteinising hormone (IU/l)	8	6.0 (5.6); 435	6.3 (5.6); 362
Sex hormone (nmol/l)	15	33.8 (16.6); 1256	32.7 (16.2); 1190
CV reported medical history			
Unspecified	1	13/45 (28.9)	5/43 (11.6)
Angina	1	5/21 (23.8)	5/19 (26.3)

TABLE 5 Baseline characteristics of the participants enrolled in the 17 studies that contributed to the IPD analysis (*continued*)

Baseline characteristics	Number of studies	TRT	Placebo
Coronary heart disease	7	95/803 (11.8)	82/771 (10.6)
Myocardial infarction	6	81/970 (8.4)	83/964 (8.6)
Arrhythmia	6	36/713 (5.1)	25/677 (3.7)
Peripheral vascular disease	4	12/500 (2.4)	9/472 (1.9)
Atherosclerosis	3	16/531 (3.0)	7/527 (1.3)
Heart failure	6	13/624 (2.1)	3/591 (0.5)
Valvular heart disease	4	2/586 (0.3)	9/55 (16.4)
Stable angina	3	4/530 (0.8)	8/533 (1.5)
Aortic aneurysm	2	2/379 (0.5)	5/376 (1.3)
Unstable angina	2	0/513 (0)	1/508 (0.2)
Cardiac arrest	1	0 (0)	1/110 (0.9)
CBV reported medical history	8	37/1139 (3.2)	58/1085 (5.4)
Diabetes ^a	12	432/1574 (27.5)	402/1492 (26.9)
Prostate cancer	17	0/1750 (0)	0/1681 (0)
QoL			
SF-36/SF-12	5		
Physical functioning		50.40 (7.81); 305	50.05 (7.96); 275
Role-physical		46.04 (13.71); 304	45.58 (14.21); 274
Bodily pain		52.46 (9.12); 299	51.23 (8.98); 272
General health		49.65 (9.85); 305	49.26 (8.96); 273
Vitality		54.80 (9.58); 305	54.34 (9.56); 275
Social functioning		51.18 (8.38); 305	51.06 (8.48); 274
Role emotional		44.42 (16.38); 304	43.52 (17.15); 275
Mental health		53.06 (8.02); 305	52.52 (8.66); 275
Physical health composite score		50.04 (8.73); 298	49.54 (7.83); 269
Mental health composite score		50.53 (11.07); 298	50.00 (11.48); 269
AMS			
Total	8	38.91 (12.36); 549	37.05 (11.42); 519
Somatic subscale	5	8.88 (4.01); 344	8.44 (3.87); 338
Psychological subscale	5	14.91 (5.35); 335	14.53 (4.86); 337
Sexual subscale	5	12.20 (4.19); 346	11.90 (4.27); 336
Sexual function			
The IIEF-15	5		
Total		33.47 (20.65); 800	31.11 (20.84); 818
Erectile function		13.12 (10.03); 814	12.02 (10.00); 838
Orgasmic function		5.28 (3.91); 820	4.76 (4.02); 841
Sexual desire		5.18 (2.12); 819	5.03 (2.12); 839

continued

TABLE 5 Baseline characteristics of the participants enrolled in the 17 studies that contributed to the IPD analysis (*continued*)

Baseline characteristics	Number of studies	TRT	Placebo
Intercourse satisfaction		5.27 (5.00); 818	4.65 (4.96); 844
Overall satisfaction		4.65 (2.48); 808	4.59 (2.52); 826
The IIEF-5	5	14.66 (7.16); 273	14.74 (7.01); 206
Androgen deficiency in ageing males	1	4.06 (2.21); 113	3.69 (2.43); 110
Physiological markers			
Testosterone (nmol/l)	16	9.21 (2.85); 1387	9.21 (2.83); 1318
Free testosterone (pmol/l)	12	196.02 (66.46); 120	198.92 (70.87); 116
Fasting glucose (mmol/l)	12	6.55 (2.18); 1421	6.66 (2.36); 1353
Cholesterol (mmol/l)	15	4.71 (1.12); 1670	4.73 (1.10); 1606
Low-density lipoproteins (mmol/l)	15	2.81 (1.02); 1644	2.78 (1.00); 1584
High-density lipoproteins (mmol/l)	15	1.20 (0.36); 1664	1.21 (0.39); 1599
Triglyceride (mmol/l)	15	1.87 (1.39); 1653	1.91 (1.50); 1584
Hb (g/l)	14	145.26 (12.64); 160	144.30 (12.89); 151
HbA1c (%)	10	6.35 (1.08); 1067	6.36 (1.12); 1059
Haematocrit (%)	16	43.29 (3.68); 1694	42.99 (3.83); 1621
SBP (mmHg)	12	133.13 (17.30); 130	133.52 (16.62); 127
DBP (mmHg)	12	77.21 (10.74); 1300	77.08 (10.72); 1274
Psychological symptoms			
BDI	3	10.01 (7.99); 158	9.36 (7.57); 113

a Type 1, 2 and unknown type.

Note

Values are mean (standard deviation); numbers or numbers (per cent).

Eight of the 18 studies reported mean baseline testosterone in units of nmol/l^{53,59,60,62,65,66,68,69} and nine studies in units of ng/dl,^{54–58,63,64,67,70} which were converted to nmol/l, as described above.

Mean testosterone in the active treatment groups ranged from 7.5⁶⁷ to 13.2 nmol/l,⁵⁸ and in the placebo groups from 7.6⁶⁷ to 14.5 nmol/l.⁵⁸ The inclusion criteria in the study by Kenny (2010) stated that ‘men were selected for testosterone levels less than 350 ng/dl or bioavailable testosterone levels at least 1.5 standard deviations (SDs) lower than those of young adult men (95–350 ng/dl for men aged 40–49).’⁵⁸ However, mean (SD) baseline testosterone levels reported were equivalent to 13.2 and 14.5 nmol/l, in the TRT and control groups, respectively. The reasons for this disparity are unclear and the study’s corresponding author did not respond to our clarification request. The remaining study did not explicitly report baseline testosterone status but stated that the mean across all groups at baseline was equivalent to < 11.1 nmol/l.⁶¹

Individual participant data integrity

For the 17 studies that provided IPD, a total of 3474 participants were randomised and, of these, 3423 participants were included in the published analyses. Differences between numbers of randomised and analysed participants were justified by the number of participants who withdrew consent, had no follow-up data, were randomised in error, or had an adverse event and were therefore not included. A total of 3423 participants were included in the published studies for which IPD was obtained; however,

we received IPD for 3431 participants. One study⁴¹ provided data for all randomised participants ($n = 44$) and not those analysed ($n = 40$) and another study⁵¹ provided data for 40 participants instead of the 35 randomised participants included in the publication. In both cases, the collaborators were unable to confirm the exact participants for whom data were sent. Another study⁴² provided 189 instead of 190 analysed, with 103 randomised to TRT and 86 to placebo instead of 92 and 98, respectively. After communicating with the collaborator, we were not able to clarify this issue. For the remaining 14 studies, all data provided were those used in the analysis of the published paper and there were no major discrepancies in outcome data.

Risk of bias within studies

All studies included in the review were assessed for risk of bias using the original version of the Cochrane Collaboration's risk of bias tool for RCTs. For the 17 studies that provided IPD, risk of bias assessment was conducted in two ways: first, using only information available in the publication(s) and second, using information in the relevant publication(s) as well as any further information provided by the trial's authors/collaborators.

Individual participant data studies: published data and additional data from collaborators

A summary of the risk of bias assessments for the 17 trials that provided IPD, including data from publications and any further data received upon request from the collaborators, is presented in [Figure 2](#). Risk of bias assessment of individual trials is presented in [Figure 3](#).

For the 'selection bias' domain, the majority of trials were assessed as being at low risk of bias,^{11,12,38-40,44-48,50-52} with four trials reporting insufficient information on which to make a judgement.^{41-43,49} For the 'performance bias' domain, two studies were rated to be at high risk of bias because the research pharmacist involved in the study was aware of the randomisation^{38,40} and four studies at unclear risk of bias because, despite being described as 'double blind', no clarification was provided on who was actually blinded.^{41-43,52} The remaining 11 trials were assessed at low risk of bias for this domain.^{11,12,39,44-51}

The 'detection bias' domain was assessed at low risk of bias for most trials.^{11,12,38-42,44-46,48-51} In three trials described as 'double blind', it was unclear whether the outcome assessor had been blinded.^{42,43,52} For 'attrition bias', one trial was judged at high risk of bias due to the high number of dropouts.¹² Two trials did not provide sufficient information on which to make a robust judgement.^{38,40} The remaining 14 trials were judged at low risk of bias for this domain. 'Reporting bias' was unclear for nine studies as it was not possible to check the respective protocols.^{38-41,47,48,50-52} One study was assessed as high risk of reporting bias as some outcomes not specified in the research protocol were reported in the full publication (e.g. aerobic performance, liver volume) while other outcomes specified in the research protocol (i.e. tests

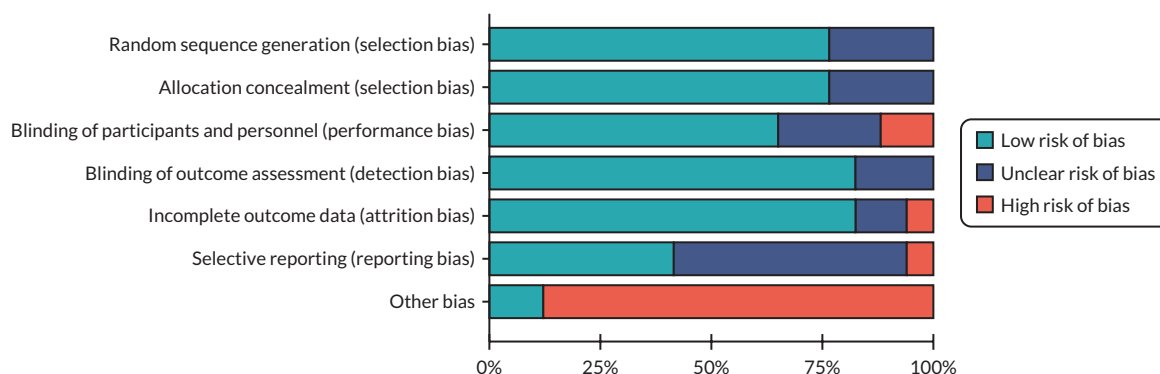


FIGURE 2 Summary of the risk of bias assessment of the 17 IPD studies (data derived from published reports and from contacting the collaborators).

	Random sequence generation (selection bias)	Allocation concealment (selection bias)	Blinding of participants and personnel (performance bias)	Blinding of outcome assessment (detection bias)	Incomplete outcome data (attrition bias)	Selective reporting (reporting bias)	Other bias
Amory 2004	+	+	-	+	?	?	+
Basaria 2010	+	+	+	+	-	-	-
Basaria 2015	+	+	+	+	+	?	-
Brock 2016	+	+	?	?	+	?	-
Emmelot-Vonk 2008	+	+	+	+	+	+	-
Gianatti 2014a	+	+	+	+	+	+	-
Giltay 2010a	+	+	+	+	+	+	-
Groti 2018	+	+	+	+	+	?	-
Hackett 2013	?	?	?	?	+	?	-
Hildreth 2013	+	+	-	+	?	?	+
Ho 2012	?	?	+	+	+	+	-
Magnussen 2016	+	+	+	+	+	?	-
Marks 2006	?	?	?	+	+	?	-
Merza 2006	?	?	?	?	+	+	-
Snyder 2016	+	+	+	+	+	+	-
Srinivas-Shankar 2010	+	+	+	+	+	+	-
Svartberg 2008	+	+	+	+	+	?	-

FIGURE 3 Risk of bias assessments of the 17 IPD studies (data derived from published reports and from contacting the collaborators).

of balance, reaction time, muscle volume, Psychological Well Being index, physical activity) were not.¹² The remaining seven studies were assessed as low risk of bias for this domain.^{11,43-47,49} For the ‘other bias’ domain, two studies were judged at low risk of bias,^{38,40} while the remaining 15 studies were judged at high risk of bias due to financial or other connections with the pharmaceutical industry.^{11,12,38,39,41-48,50-52} Overall risk of bias was judged to be unclear for 5 studies^{41-43,49,52} and low for 12 studies.^{11,12,38-40,44-48,51}

Individual participant data studies: published data only

Risk of bias of the 17 studies included in the IPD analysis was also conducted, based on data available in the relevant publication(s) only. A summary of the risk of bias assessments is presented in [Figure 4](#). Risk of bias of individual trials is presented in [Figure 5](#).

In general, risk of bias assessment of published studies was not hugely dissimilar to that conducted after the IPD were obtained. In particular, further information obtained by the authors of six studies allowed us to re-classify a total of 15 domains originally assessed as 'unclear' risk of bias.⁴⁵⁻⁵¹ In all cases, the judgement was changed to low risk of bias. The most common domain for re-classification was 'detection bias', in which the blinding of the outcome assessor was unclear in the original publication but subsequently clarified by the respective authors.^{45,48-51} Blinding of participants/personnel was confirmed by three authors,^{45,48,51} as well as random sequence generation and allocation concealment.^{47,48,51}

Non-IPD studies

A summary of the risk of bias assessments for the 18 trials that did not provide IPD is presented in [Figure 6](#). Risk of bias of individual trials is presented in [Figure 7](#).

For the 'selection bias' domain, six studies were judged at low risk of bias for random sequence generation,^{53,54,56,63,68,70} with four of these six also being at low risk of bias for allocation concealment.^{54,63,68,70} The remaining studies did not report sufficient information on which to base a robust judgement.^{55,57-62,64-67,69}

Three studies were judged at high risk of bias for 'performance bias': one study involved three randomised groups, of which one was open label and two were double blinded;⁵⁹ one study was open label;⁶⁷ and in a third study the pharmacist was not blinded.⁶⁶ Nine studies were judged to be at low risk of bias^{53,54,56-58,63,65,69,70} and the remaining six studies did not provide sufficient information on which to make a robust judgement.^{55,60-62,64,68} Two studies were judged at high risk of bias for 'detection bias': one study involved three randomised groups, of which one was open label and two were double blinded⁵⁹ and one study was open label.⁶⁷ Seven studies were judged to be at low risk in this domain^{53,54,56,58,63,69,70} and nine studies did not report sufficient information on which to base a robust judgement.^{55,57,60-62,64-66,68} Eight studies were judged to be at high risk of bias for 'attrition bias' due to the numbers of dropouts.^{53,55-58,64,68,69} Six studies were judged to be at low risk of bias^{54,60,63,65,66,70} and four studies did not provide sufficient information on which to base a robust judgement.^{59,61,62,67} Two studies were judged to be at high risk of 'reporting bias': in one study, some outcomes that were assessed were not reported, for example, liver and kidney functions serum bilirubin, gamma-glutamyl transferase, serum glutamate oxaloacetate transaminase and serum glutamate pyruvate transaminase, albumin, and creatinine. Other outcomes were reported but were not specified as being assessed, for

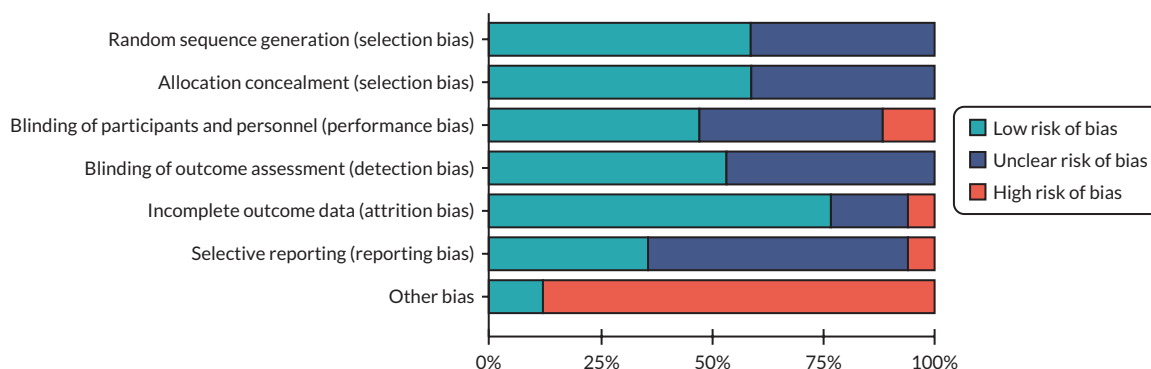


FIGURE 4 Summary of the risk of bias assessment of the 17 IPD studies (data from published reports only).

	Random sequence generation (selection bias)	Allocation concealment (selection bias)	Blinding of participants and personnel (performance bias)	Blinding of outcome assessment (detection bias)	Incomplete outcome data (attrition bias)	Selective reporting (reporting bias)	Other bias
Amory 2004	+	+	-	+	?	?	+
Basaria 2010	+	+	+	+	-	-	-
Basaria 2015	+	+	+	+	+	?	-
Brock 2016	+	+	?	?	+	?	-
Emmelot-Vonk 2008	+	+	?	?	+	+	-
Gianatti 2014a	+	+	+	+	+	?	-
Giltay 2010a	?	?	+	+	+	+	-
Groti 2018	?	?	?	?	?	?	-
Hackett 2013	?	?	?	?	+	?	-
Hildreth 2013	+	+	-	+	?	?	+
Ho 2012	?	?	+	?	+	+	-
Magnussen 2016	+	+	+	?	+	?	-
Marks 2006	?	?	?	+	+	?	-
Merza 2006	?	?	?	?	+	+	-
Snyder 2016	+	+	+	+	+	+	-
Srinivas-Shankar 2010	+	+	+	+	+	+	-
Svartberg 2008	?	?	?	?	+	?	-

FIGURE 5 Risk of bias assessments of the 17 IPD studies (data from publications only).

example, QUICKI index.⁶⁰ Similarly, blood pressure, urinalysis and urine flow rates were assessed but not reported in another study.⁶⁵ Reporting bias was unclear for the remaining 16 studies as it was not possible to check the respective protocols.^{53-59,61-64,67-70} Three studies were judged to be at low risk of bias for 'other bias',^{60,63,67} three studies did not report sufficient information on which to make a robust judgement,^{61,62,65} and 12 studies were judged at high risk of bias due to financial or other connections

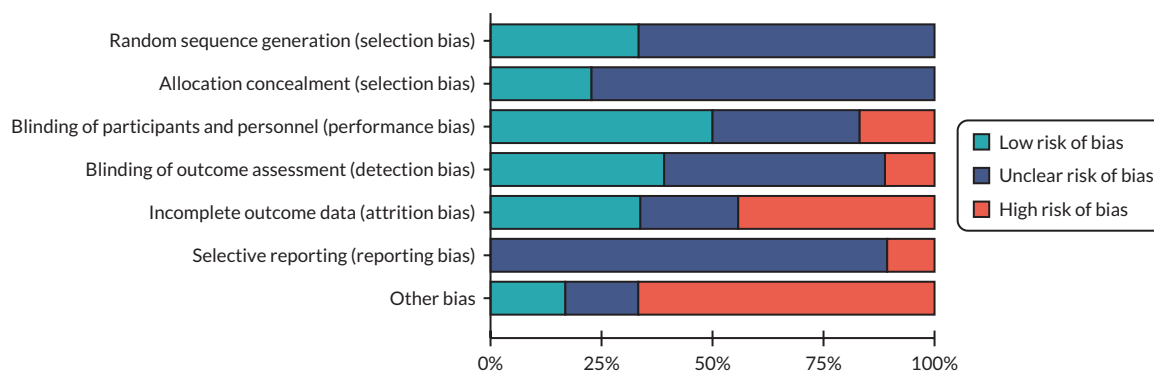


FIGURE 6 Summary of risk of bias assessment of the 18 non-IPD studies.

with the pharmaceutical industry.^{53-59,64,66,68-70} Overall risk of bias was judged to be high for two studies,^{59,67} unclear for 13 studies,^{53,55-58,60-62,64-66,68,69} and low for three studies.^{54,63,70}

In general, a comparison of the risk of bias assessments of the 17 IPD studies and the 18 non-IPD studies showed a greater proportion of low risk of bias assessments for the IPD studies. However, the higher proportion of unclear risk of bias assessments of the non-IPD studies makes for an imbalanced comparison, and no robust conclusions can be drawn regarding the relative risk (RR) of bias across the two classes of studies.

Primary outcomes

All-cause mortality

Overall effect

Based on the one-stage fixed-effects IPD meta-analysis, mortality from any cause (14 studies, 3158 men) was lower among participants treated with TRT [6/1621 (0.4%)] than placebo [12/1537 (0.8%)], but the 95% CI was wide (OR 0.46, 95% CI 0.17 to 1.24; p -value 0.13; [Table 6](#)). Causes of death included myocardial infarction, cancer and ruptured aortic aneurysm. Based on data we received, we could not determine the cause of eight deaths from three studies.

For the two-stage analysis, the IPD studies and aggregate data gave a similar combined effect of OR 0.63, 95% CI 0.30 to 1.32; $I^2 = 0\%$; however, the 95% CIs were wide for each of the individual trials for both the IPD and non-IPD analyses (see [Figure 10a](#)). Details of the causes of death per treatment groups are shown in [Appendix 6, Table 38](#). Sensitivity analysis using Mantel-Haenszel showed similar results (see [Figure 8b](#)).

Publication bias

The contour-enhanced funnel plots as well as Peter's test on small-study effects for IPD (Peter's test $p = 0.283$), aggregate and all studies combined (Peter's test $p = 0.458$) show no evidence of significant small-study bias (see [Appendix 6, Figure 21](#)).

Cardiovascular and/or cerebrovascular events

Overall effect

In total, 13 studies with 3120 men reported whether CV and/or CBV events occurred within 12 months (see [Table 7](#)). There were 120/1601 (7.5%) participants in the TRT group and 110/1519 (7.2%) participants in the placebo group with no evidence of a difference between the two treatment groups (OR 1.07, 95% CI 0.81 to 1.42; $p = 0.62$). Results were not changed significantly by including follow-up

time as a weight in the model (data not shown). Most of the participants had a CV rather than a CBV event, with some participants experiencing more than one event. The most frequent events were the following: arrhythmia, coronary heart disease (CHD), heart failure, myocardial infarction and valvular heart disease. The sensitivity analysis of including unknown cause of death did not change the results (OR 1.05, 95% CI 0.79 to 1.38; $p = 0.740$).

	Random sequence generation (selection bias)	Allocation concealment (selection bias)	Blinding of participants and personnel (performance bias)	Blinding of outcome assessment (detection bias)	Incomplete outcome data (attrition bias)	Selective reporting (reporting bias)	Other bias
Aversa 2010a	?	?	?	?	+	-	+
Aversa 2010b	?	?	?	?	?	?	?
Basurto 2008	+	+	+	+	+	?	+
Behre 2012	+	+	?	?	-	?	-
Borst 2014	+	?	+	+	-	?	-
Cavallini 2004	?	?	?	?	?	?	?
Cherrier 2015	+	+	+	+	+	?	-
Chiang 2007	?	?	?	?	-	?	-
Clague 1999	?	?	+	?	+	-	?
Dhindsa 2016a	?	?	?	?	+	?	-
Dias 2016	+	?	+	+	-	?	-
Jones 2011	?	?	+	+	-	?	-
Kaufman 2011a	?	?	+	?	-	?	-
Kenny 2010	?	?	+	+	-	?	-
Morales 2009	?	?	-	?	+	?	-
Paduch 2015a	+	+	+	+	+	?	-
Steidle 2003	+	+	-	-	?	?	-
Wang 2013	?	?	-	-	?	?	+

FIGURE 7 Risk of bias assessments of the 18 non-IPD studies.

TABLE 6 One-stage IPD meta-analysis for mortality from any cause

Outcome	Number of studies	TRT n/N (%)	Placebo n/N (%)	OR 95% CI	p-value
Mortality from any cause ^a	14	6/1621 (0.4)	12/1537 (0.8)	0.46 (0.17 to 1.24)	0.13
Details		N = 6	N = 12		
Myocardial infarction	3	2 (33.3)	2 (16.7)		
Cancer	1	0 (0)	3 (25.0)		
Ruptured aortic aneurysm	1	0 (0)	1 (8.3)		
Constrictive pericarditis	1	1 (16.7)	0 (0)		
Multiple organ failure	1	1 (16.7)	0 (0)		
Venous thromboembolism	1	0 (0)	1 (8.3)		
Unknown	3	2 (33.3)	5 (41.7)		

a Of the 14 studies, 8 reported no deaths^{38,40,43,45,46,48,50,52} and 6 reported deaths.^{11,12,39,44,47,49} For the remaining three studies we were unable to confirm whether any deaths occurred, therefore they were not included.^{41,42,51}

Note

Values are numbers (per cent) or numbers.

For the two-stage analysis (see [Figure 9](#)), the effect estimates differed between the IPD studies (OR 1.03, 95% CI 0.77 to 1.38) and the aggregate data (OR 0.35, 95% CI 0.12 to 1.01). However, once combined, there was no evidence of a difference between the two treatment groups (OR 0.96, 95% CI 0.72 to 1.27). Further details are shown in [Appendix 6, Table 37](#).

Subgroup analyses

[Figure 10](#) shows the pre-specified subgroup analyses (assessed at a stricter significance of 1%). Overall, there was no evidence of treatment-covariate interaction for any of the pre-specified subgroup analyses [diabetes (yes/no), smoking status (yes/no), testosterone (nmol/l) and free testosterone (pmol/l)] or for the post hoc analysis [age (years)].

Publication bias

We found no visual and statistical evidence of a significant small-study bias in the contour-enhanced funnel plots and the results of the Peter's test for IPD (Peter's test $p = 0.815$), aggregate and all studies combined (Peter's test $p = 0.701$; see [Appendix 6, Figure 22](#)).

Secondary outcomes**Quality of life**

Based on the one-stage meta-analysis, the SF-36/SF-12 norm-based scores (five studies) showed evidence of an overall difference in favour of TRT compared to placebo for three of the 10 subscales: the social functioning subscale (MD 1.74, 95% CI 0.14 to 3.34; $p = 0.034$), the role emotional subscale (MD 1.66, 95% CI 0.57 to 2.76; p -value 0.003), and the mental health composite score (MD 1.95, 95% CI 0.64 to 3.26; $p = 0.004$; see [Table 8](#)). There was little evidence of heterogeneity ($\tau^2 = 1.51$, 0 and 0.20, respectively). For the remainder of the subscales, the MD favoured TRT but the 95% CIs were wide. The two-stage meta-analysis (see [Appendix 6, Figure 23a-j](#)) showed similar results but with varying heterogeneity (I^2 ranged from 0% to 68%).

Seven studies (938 men in total) provided IPD for AMS total score. The one-stage analysis (see [Table 8](#)) showed evidence of a difference in favour of TRT (MD -2.62, 95% CI -4.02 to -1.23; $p < 0.001$, $\tau^2 = 1.52$). Of these seven studies, only five provided IPD for each of the AMS subscales, but all showed

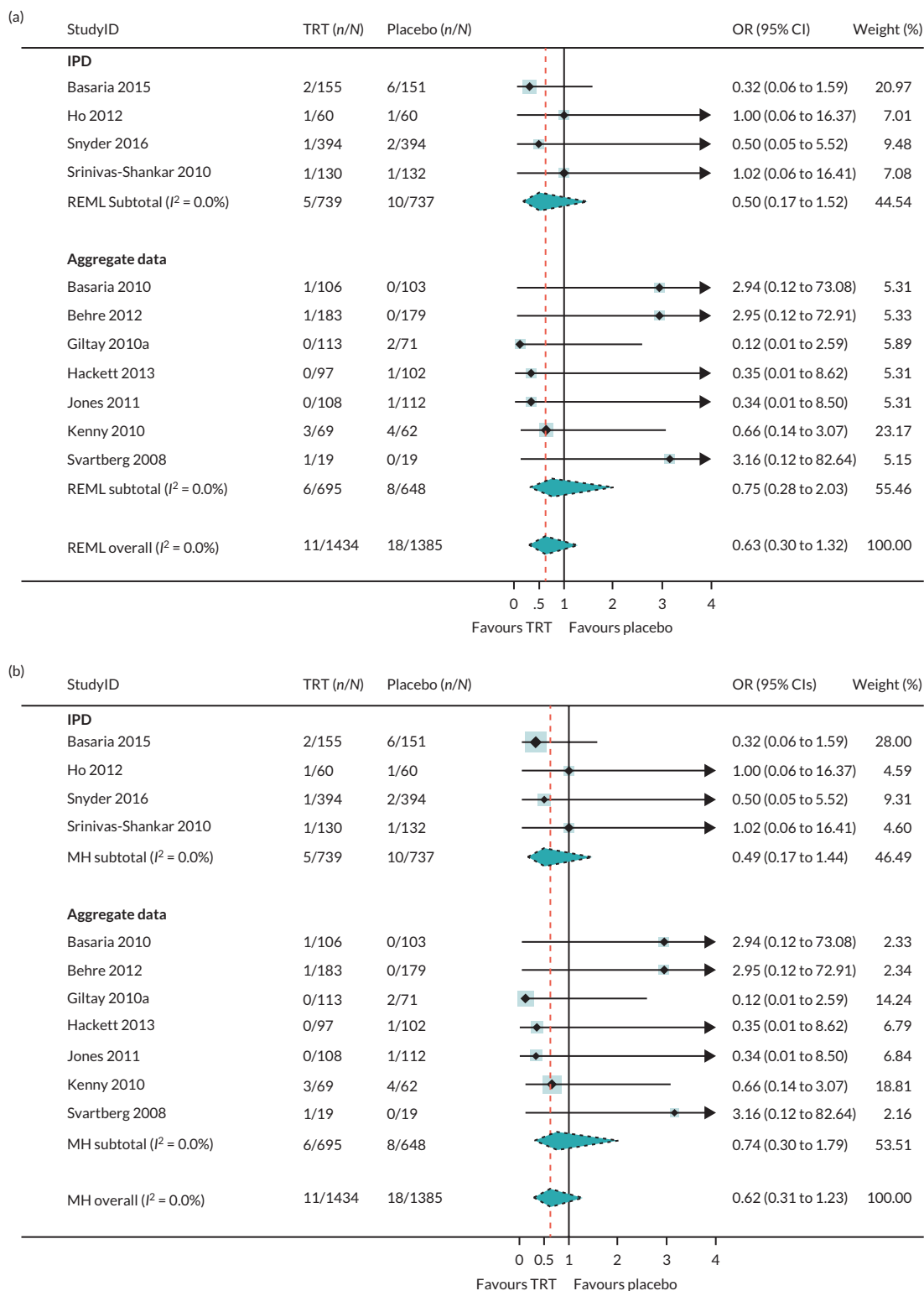


FIGURE 8 IPD meta-analysis for mortality from any cause; (a) two-stage analysis; (b) Mantel-Haenszel. MH, Mantel-Haenszel.

results in favour of TRT. For the two-stage analyses, four additional studies provided aggregate data for AMS total score but only two for AMS subscales (see Figure 11a-d). Overall, the results for the two-stage IPD analysis were consistent with the one-stage analysis although more heterogeneous. However, the aggregated data for AMS total score (and to some extent for the somatic and sexual subscale too) showed considerable heterogeneity compared to the IPD studies, with varying 95% CI widths, and

TABLE 7 One-stage IPD meta-analysis for CV and/or CBV events

Outcome	Number of studies	TRT n/N (%)	Placebo n/N (%)	OR 95% CI	p-value
Number of participants with a CV and/or CBV events ^a	13	120/1601 (7.5)	110/1519 (7.2)	1.07 (0.81 to 1.42)	0.62
Total number of a CV and/or CBV events	13	182	183		
Number of participants with a CV event	11	107/120 (89.2)	105/110 (95.5)		
Total number of CV events ^b	11	166	176		
Details					
Arrhythmia	6	52	47		
Coronary heart disease	6	33	33		
Heart failure	6	22	28		
Myocardial infarction	7	10	16		
Valvular heart disease	2	18	12		
Peripheral vascular disease	4	8	14		
Stable angina	5	7	7		
Aortic aneurysm ^c	5	6	7		
New angina	3	5	5		
Unstable angina	3	2	4		
Aortic dissection	1	2	0		
Atherosclerosis	1	1	1		
Cardiac arrest	2	0	2		
Number of participants with a CBV event	11	15/120 (12.5)	7/110 (6.4)		
Total number of CBV events ^c	11	16	7		

a Of the 13 studies, 2 reported no CV and/or CBV events.^{48,50} For the remaining four studies we were unable to confirm whether CV and/or CBV events occurred.^{41-43,51}

b Some participants had more than one event.

c One event was a ruptured aortic aneurysm.

Note

Values are numbers (per cent) or numbers.

with one study⁶¹ showing exaggerated benefit of TRT. Once combined in the final stage, this effect was dampened but the heterogeneity remained high.

Other QoL outcomes are presented in [Appendix 6, Table 40](#).

Sexual function

Five studies (1412 men in total) provided IPD for the assessment of sexual function using the IIEF-15. The one-stage analysis (see [Table 9](#)) for total score showed that compared to placebo, TRT improved sexual function (MD 5.52, 95% CI 3.95 to 7.10; p -value < 0.001, $\tau^2 = 1.17$). For the individual IIEF-15 subscales, results were similar, showing evidence of a difference in favour of TRT. The two-stage

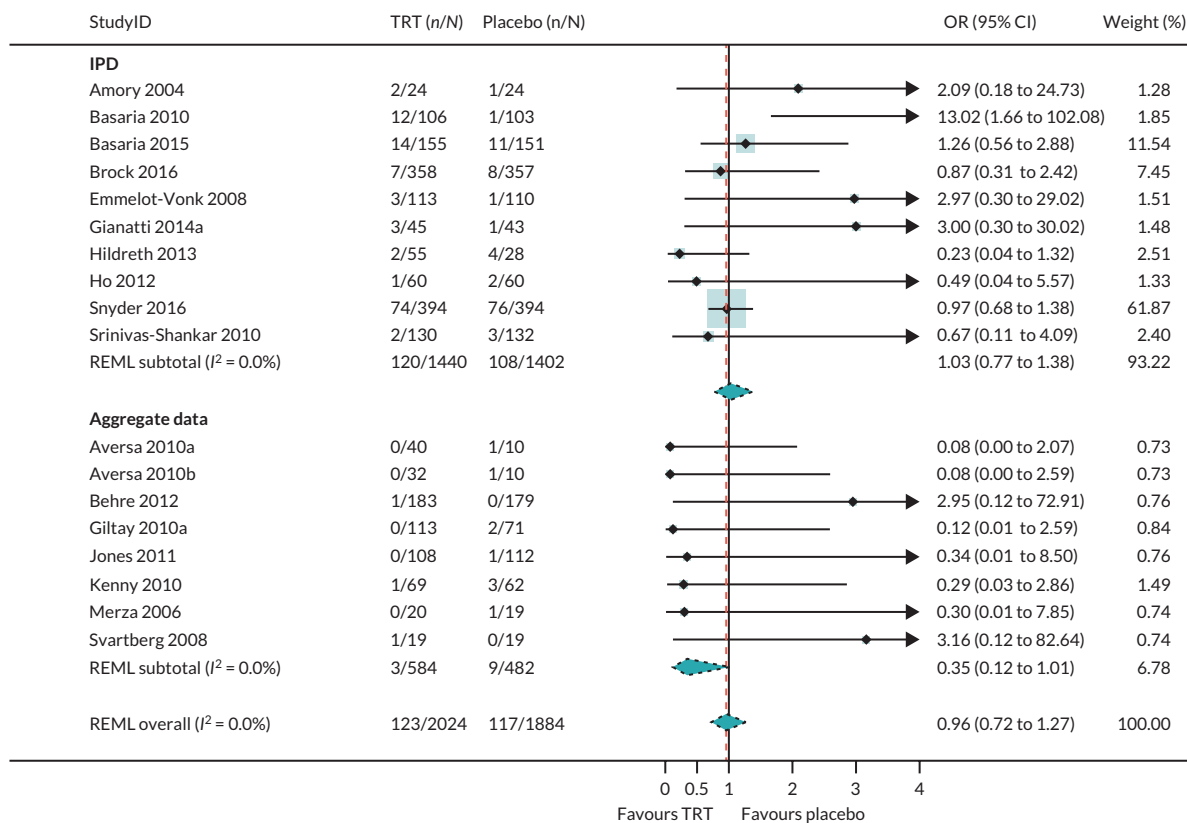


FIGURE 9 Two-stage IPD meta-analysis for CV and/or CBV events. Note: Giltay 2010a and Basaria 2010 did not converge so a fixed-effects analysis was included in the non-IPD trials.

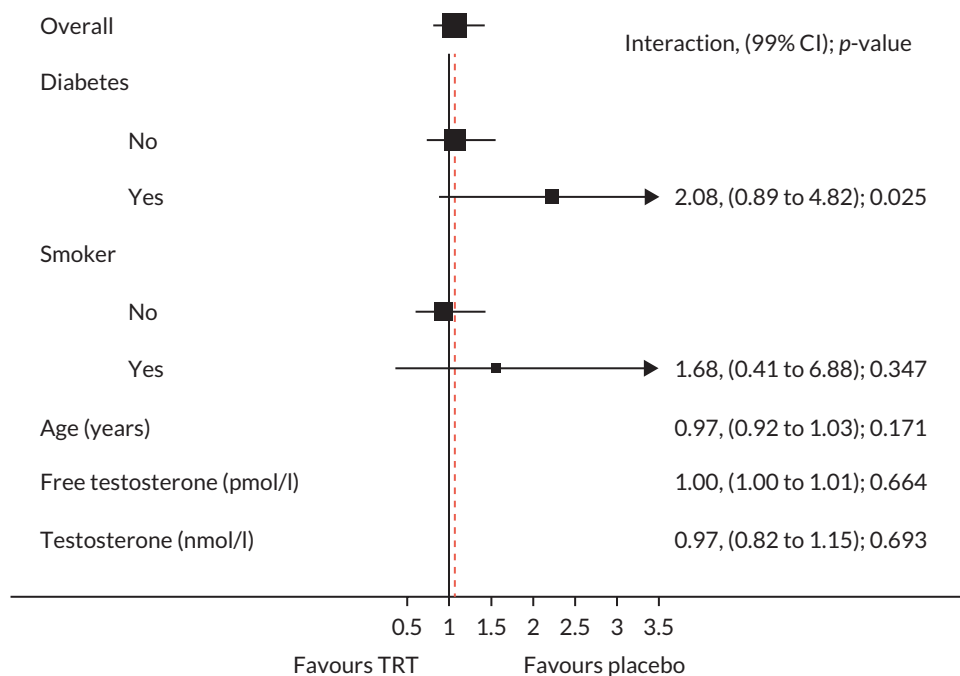


FIGURE 10 Subgroups for TRT vs. placebo.

analysis used four additional studies of aggregate data for IIEF-15 total score, erectile function, orgasmic function and sexual desire, and an additional three studies with aggregate data for intercourse satisfaction and two additional studies for overall satisfaction (see [Figure 12a-f](#)). In general, the two-stage analysis showed similar but more heterogeneous results (I^2 ranged from 0% to 97%). However,

TABLE 8 One-stage analysis for QoL outcomes

Outcome	Number of studies	TRT mean (SD); n	Placebo mean (SD); n	MD	95% CI	τ^2
SF-36/SF-12						
Norm-based scores						
Physical functioning	5	51.03 (7.49); 277	49.84 (8.25); 262	0.56	(-0.33 to 1.44)	0.00
Role-physical	5	45.64 (13.90); 277	45.13 (14.46); 261	0.72	(-0.70 to 2.14)	0.64
Bodily pain	5	52.92 (9.01); 277	51.87 (9.86); 262	0.05	(-1.17 to 1.27)	0.00
General health	5	50.46 (8.70); 277	49.71 (9.31); 262	0.77	(-1.11 to 2.65)	2.70
Vitality	5	56.67 (8.81); 277	54.55 (9.34); 263	1.78	(-0.43 to 3.99)	4.01
Social functioning	5	52.23 (7.34); 274	50.68 (8.64); 262	1.74	(0.14 to 3.34)	1.51
Role emotional	5	44.46 (15.91); 277	42.92 (17.23); 260	1.66	(0.57 to 2.76)	0.00
Mental health	5	53.88 (7.51); 277	53.14 (9.17); 263	0.41	(-0.65 to 1.47)	0.00
Physical health composite score	5	50.35 (8.08); 274	49.70 (8.47); 258	0.01	(-1.48 to 1.51)	1.36
Mental health composite score	5	51.65 (9.50); 274	50.20 (11.76); 258	1.95	(0.64 to 3.26)	0.20
AMS						
Total	7	32.19 (10.23); 482	34.22 (11.10); 456	-2.62	(-4.02 to -1.23)	1.52
Somatic subscale	5	12.73 (4.21); 315	13.25 (4.58); 307	-0.64	(-1.18 to -0.09)	0.03
Psychological subscale	5	7.71 (3.16); 309	7.99 (3.47); 312	-0.40	(-0.76 to -0.05)	0.00
Sexual subscale	5	10.33 (3.82); 320	11.12 (4.20); 324	-0.78	(-1.33 to -0.24)	0.07

Note

Outcomes were analysed using a random-effects model.

there were some differences between IPD and aggregate data with varying 95% CI widths for some of the scores.

Sexual function was also reported by a further five IPD studies (442 men in total) using the IIEF-5, a shorter version of the IIEF-15. These studies showed no evidence of difference between treatment groups (see [Table 9](#)). No aggregate data provided further data and the two-stage analysis of the IPD studies showed similar but more heterogeneous results ($I^2 = 75.6\%$) with wider 95% CIs (see [Appendix 6, Figure 24](#)).

Another sexual function scale, the ADAM, was reported by one IPD study (221 men in total). There was no evidence of a difference between treatment groups (MD -0.25, 95% CI -0.73 to 0.23; $p = 0.308$; [Appendix 6, Table 41](#)). Other sexual function outcomes reported by studies that provided IPD are presented in [Appendix 6, Table 41](#).

Two-stage meta-analyses of the remaining outcomes are presented in [Appendix 6, Figures 25–28](#). One study provided aggregate data for HED and Sexual Arousal, Interest and Drive Scale: for both

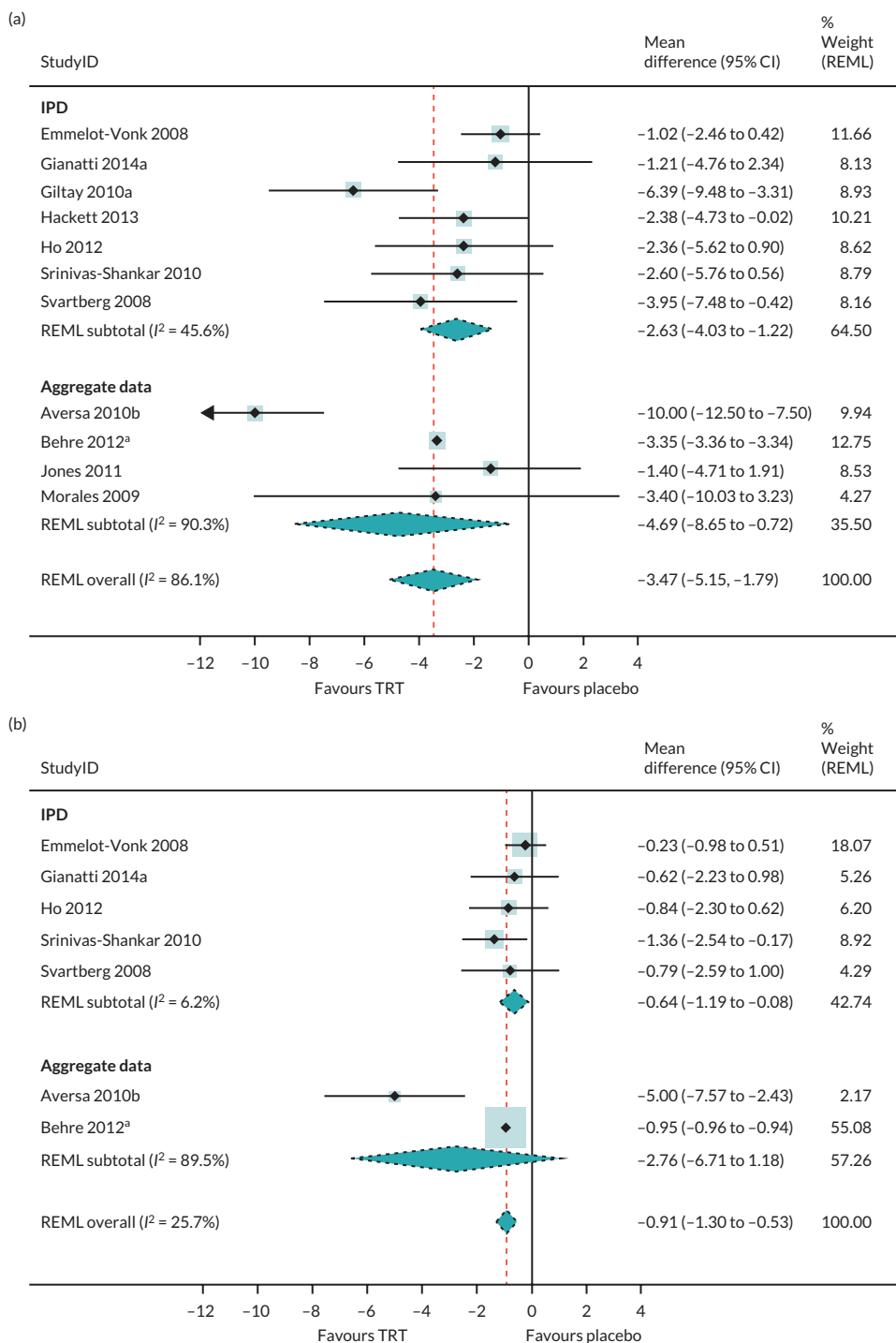


FIGURE 11 Two-stage meta-analysis for AMS. (a) Total score; (b) Somatic subscale; (c) Psychological subscale; (d) Sexual subscale. a, Data presented as change from baseline. (continued)

instruments there was evidence of a difference in favour of TRT compared to placebo (MD 3.20, 95% CI 0.84 to 5.56 and MD 4.30, 95% CI 1.60 to 7.00, respectively).

Regarding the post hoc subgroup analysis, we did not find any treatment-modifying factors (see [Appendix 6, Table 42](#)). For the threshold analysis, we identified the following thresholds for age: 52, 70, 72 and 72.8 years, which were simplified into the following age categories < 52, 52–70 and > 70 (see [Appendix 6, Table 43](#) and [Appendix 6, Figure 29a](#)). For total serum testosterone, a threshold was found at 9.8 nmol/l and 30.6 for BMI (see [Appendix 6, Table 43](#) and [Appendix 6, Figures 29](#) and [30](#)).

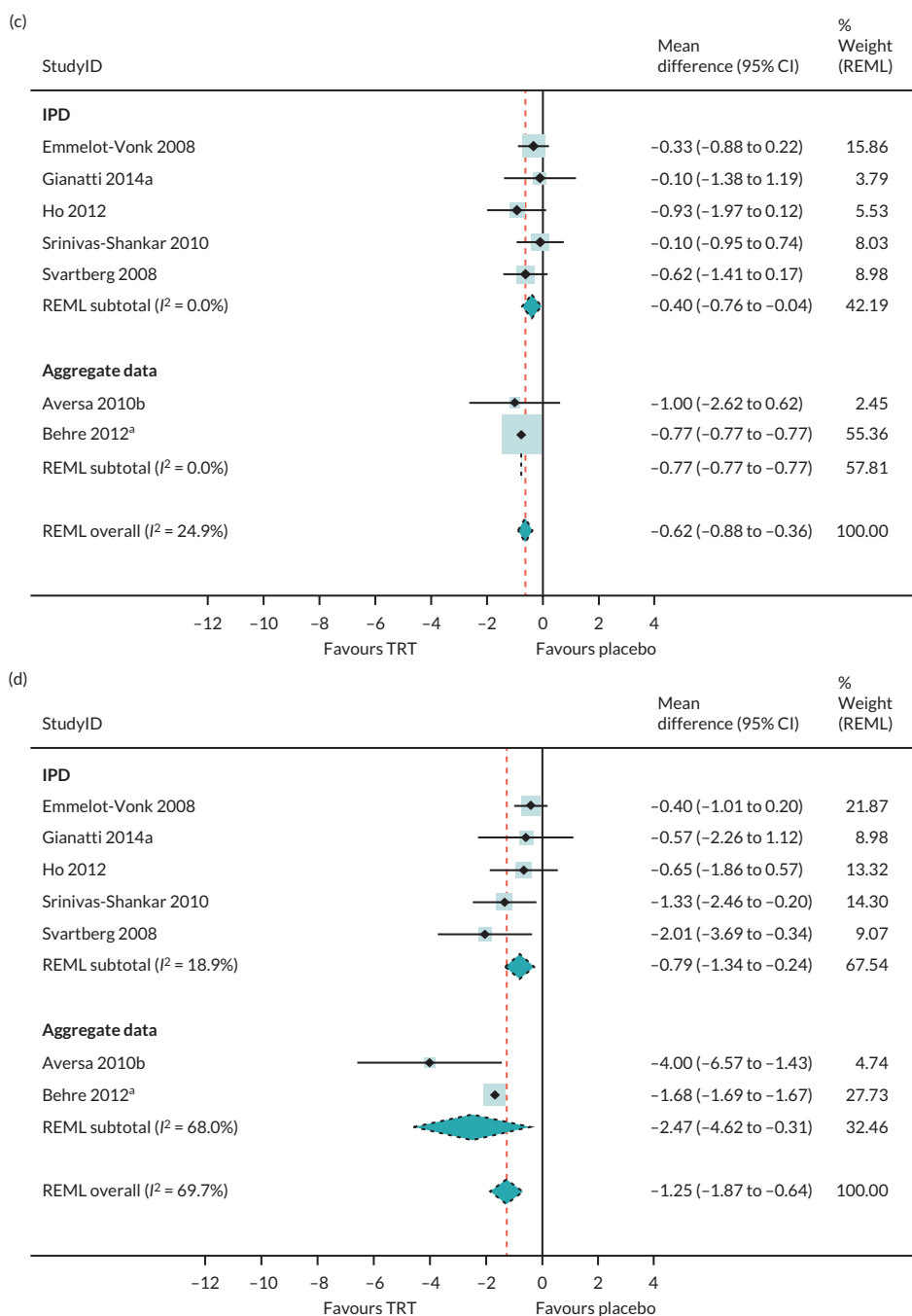


FIGURE 11 Two-stage meta-analysis for AMS. (a) Total score; (b) Somatic subscale; (c) Psychological subscale; (d) Sexual subscale. a, Data presented as change from baseline.

Physiological markers

For the one-stage meta-analysis for level of testosterone (nmol/l), 16 studies (2308 men in total) provided IPD and showed evidence of higher testosterone levels in the treatment group compared to the placebo group (MD 7.24, 95% CI 5.07 to 9.41); p -value < 0.001, $\tau^2 = 17.01$). Similar findings were observed for free testosterone (pmol/l) but with substantial heterogeneity (MD 186.40, 95% CI 115.91 to 256.90; p < 0.001, $\tau^2 = 13741.90$). For both cholesterol (mmol/l) and triglycerides (mmol/l), there was evidence of some difference and a degree of homogeneity (MD -0.15, 95% CI -0.20 to -0.10; p < 0.001, $\tau^2 = 0.00$ and MD -0.09, 95% CI -0.18 to -0.00; p = 0.044, $\tau^2 = 0.01$, respectively). Similar results were observed for Hb and haematocrit. For HbA1c (%) and blood pressure, there was no evidence of a difference between treatment groups (see [Table 10](#)).

TABLE 9 One-stage analysis for sexual function outcomes

Outcome	Number of studies	TRT mean (SD); n	Placebo mean (SD); n	MD	95% CI	τ^2
The IIEF-15						
Total	5	40.67 (21.51); 703	33.77 (22.44); 709	5.52	(3.95 to 7.10)	1.17
Erectile function	5	15.98 (10.32); 714	13.15 (10.62); 722	2.14	(1.40 to 2.89)	0.64
Orgasmic function	5	6.11 (3.78); 714	5.08 (4.14); 726	0.81	(0.48 to 1.14)	0.27
Sexual desire	5	6.04 (2.15); 716	5.21 (2.25); 724	0.80	(0.62 to 0.97)	0.00
Intercourse satisfaction	5	6.67 (5.19); 714	5.01 (5.17); 725	1.33	(0.95 to 1.71)	0.15
Overall satisfaction	5	5.70 (2.66); 706	5.10 (2.66); 711	0.52	(0.29 to 0.74)	0.02
The IIEF-5	5	16.73 (6.94); 251	15.90 (7.16); 191	0.22	(-1.64 to 2.08)	5.19

Notes

Values are mean (SD); numbers.
Outcomes were analysed using random effects.

The results of one-stage meta-analyses for other physiological markers are shown in [Appendix 6, Table 44](#).

The two-stage meta-analysis (see [Appendix 6, Figures 31–43](#)) showed similar results, but some of the individual studies had wide 95% CIs with varying heterogeneity. There were also some differences between IPD studies and aggregate data.

Psychological symptoms

Individual participant data for BDI was provided by three studies (246 men in total; see [Appendix 6, Table 45](#)). The one-stage analysis showed no evidence of a difference between the treatment groups (MD -1.10, 95% CI -2.49 to 0.30; $p = 0.123$, $\tau^2 = 0.71$). Aggregate studies provided no further data. Regarding the remaining psychological outcomes, only one study provided IPD (see [Appendix 6, Table 45](#)). The two-stage meta-analyses are presented in [Appendix 6, Figures 44–47](#).

Post hoc subgroup analyses

[Figure 13](#) shows the post hoc subgroup analyses for the following outcomes: AMS, IIEF-15, IIEF-5, Hb (g/l) and BDI. Overall, there was no evidence of consistent treatment-covariate interaction between the effect of TRT and the considered subgroups [diabetes (yes/no), smoking status (yes/no), age (years), testosterone (nmol/l) and free testosterone (pmol/l)] for any of the assessed outcomes.

Additional outcomes

Two IPD studies reporting diabetes complications showed no evidence of difference between the treatment groups (see [Table 11](#)). Similarly, no evidence of a difference between treatment groups was found for hypertension, venous thromboembolism or non-stroke CBV (see [Table 11](#)). Prostate cancer, which was assessed by four IPD studies and four aggregate data, also indicated no evidence of difference. For both oedema and high haematocrit reported by the IPD studies, there was evidence of raised levels in the TRT group compared to the placebo group.

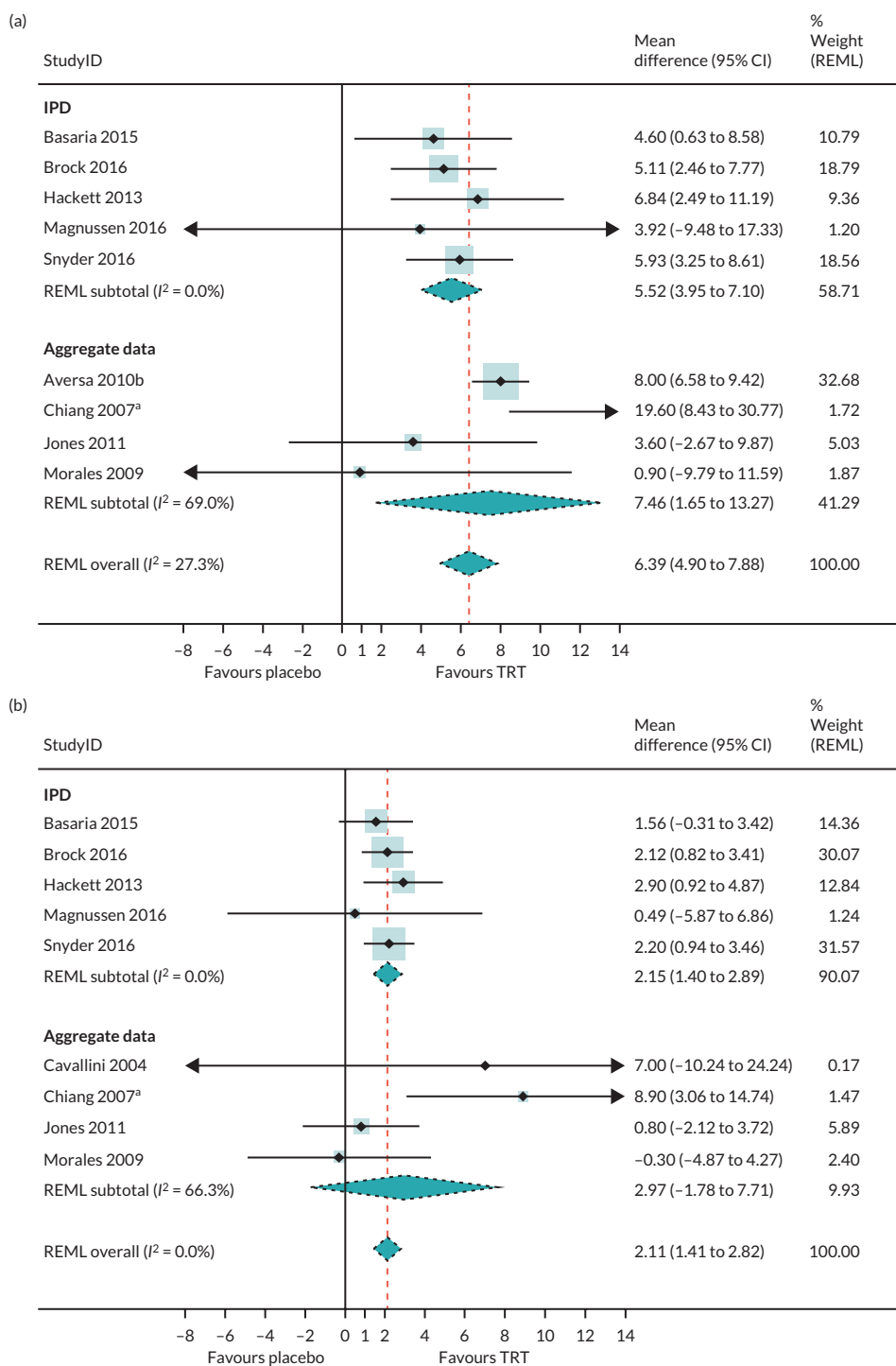


FIGURE 12 Two-stage meta-analysis for IIEF-15. (a) Total score; (b) Erectile function; (c) Orgasmic function; (d) Sexual desire; (e) Intercourse satisfaction; (f) Overall satisfaction. a, Data presented as change from baseline. (continued)

Discussion

Published systematic reviews and meta-analyses have provided conflicting and inconclusive results on the CV risk of TRT; these studies have been heterogeneous in terms of inclusion criteria, and the definition and choice of outcome measures.¹⁷⁻²⁶ Moreover, most of these reviews included participants with different baseline levels of low testosterone, which is an accepted proxy for disease severity in

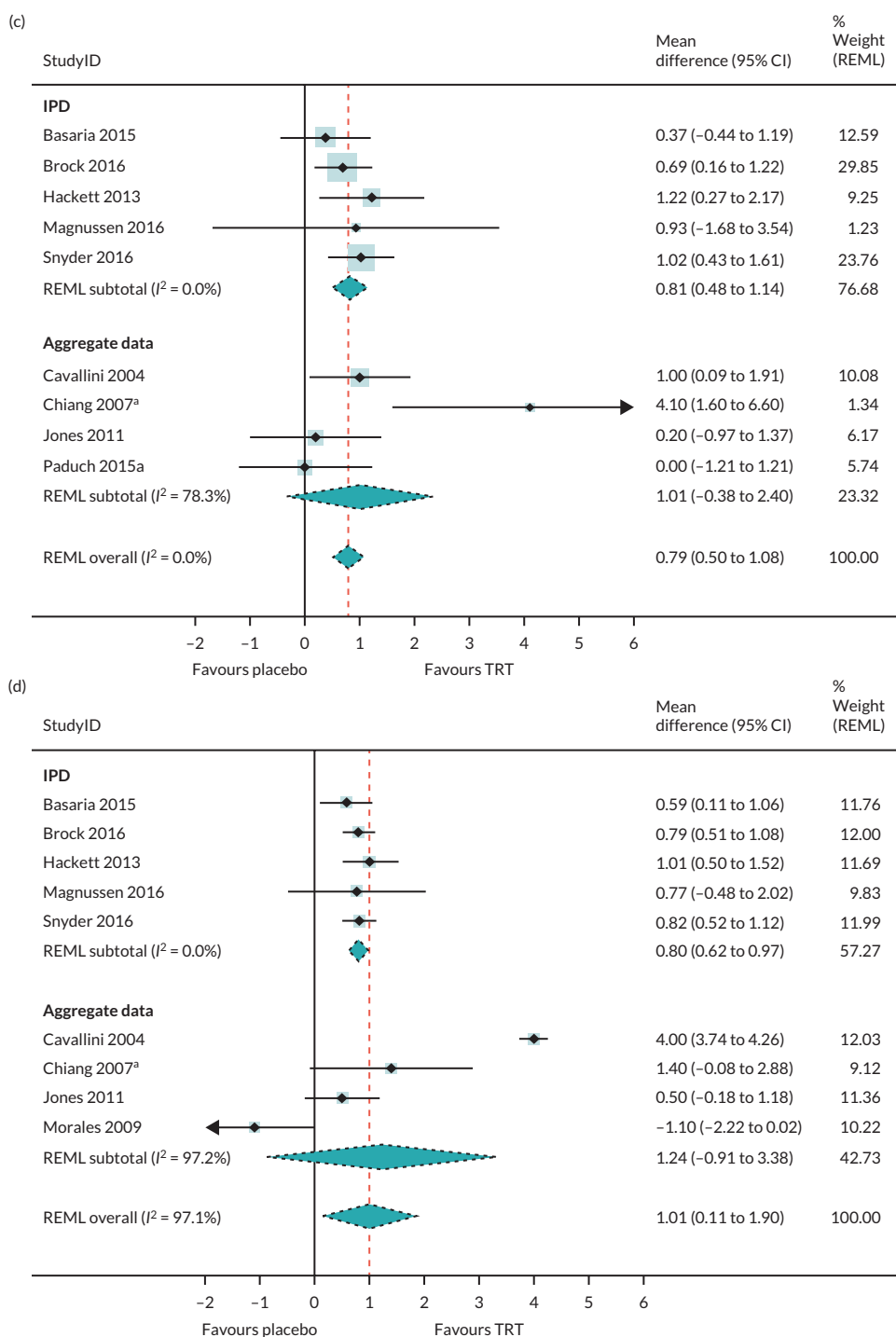


FIGURE 12 Two-stage meta-analysis for IIEF-15. (a) Total score; (b) Erectile function; (c) Orgasmic function; (d) Sexual desire; (e) Intercourse satisfaction; (f) Overall satisfaction. a, Data presented as change from baseline. (continued)

MH.⁴ To overcome these limitations, this IPD meta-analysis has allowed us to confirm the integrity and classification of data, while evaluating which factors may influence the safety and efficacy of TRT. We have observed that the recorded frequency of CV events during RCTs is similar with TRT compared to placebo. We have also determined that TRT did not increase the risk of any observed subtype of CV event. In line with the literature, TRT improved sexual function and QoL.⁷² There were no significant effects of TRT on blood pressure, serum lipids or glycaemic markers when compared with placebo in men with MH. Testosterone stimulates haematopoiesis, so it is unsurprising that the risk of polycythaemia was significantly elevated by TRT in men with hypogonadism.⁷³ However, TRT did not increase deep vein thrombosis (DVT) risk.

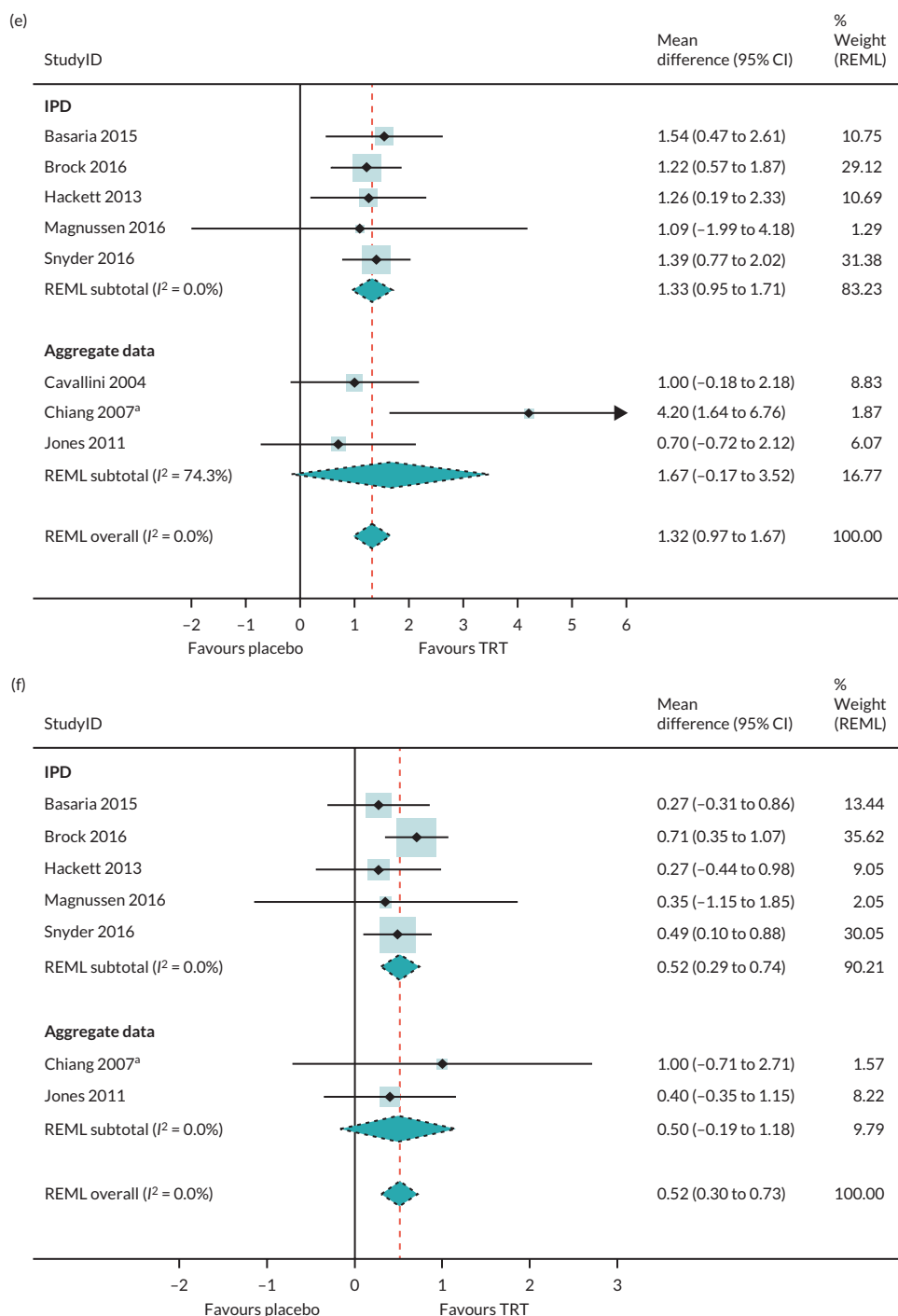


FIGURE 12 Two-stage meta-analysis for IIEF-15. (a) Total score; (b) Erectile function; (c) Orgasmic function; (d) Sexual desire; (e) Intercourse satisfaction; (f) Overall satisfaction. a, Data presented as change from baseline.

A strength of this IPD meta-analysis was the ability to perform relevant subgroup analyses. Neither patient age nor serum TT at baseline was associated with increased risk of major adverse cardiovascular events (MACE) during TRT. Two previous studies have reported that TRT is associated with lower mortality in men with both MH and type 2 diabetes.^{74,75} Our IPD meta-analysis reported a non-significant increase in MACE risk during TRT in men with MH and type 2 diabetes; this suggests that there may be heterogeneity in the risk of MACE events in this patient subgroup. Therefore, there exists insufficient evidence to conclude that men with type 2 diabetes have altered risks of TRT when compared with men without diabetes. The small total number of deaths within the IPD analysis

TABLE 10 One-stage meta-analysis for physiological marker outcomes

Outcome	Number of studies	TRT mean (SD); n	Placebo mean (SD); n	MD	95% CI	τ^2
Testosterone (nmol/l)	16	17.27 (10.34); 1211	9.87 (3.98); 1156	7.24	(5.07 to 9.41)	17.01
Free testosterone (pmol/l)	12	426.70 (368.42); 1058	203.57 (86.24); 1027	186.40	(115.91 to 256.90)	13,741.90
Fasting glucose (mmol/l)	12	6.50 (2.09); 1259	6.75 (2.38); 1181	-0.16	(-0.24 to -0.07)	0.00
Fasting glucose (mmol/l) sensitivity ^a	11	6.04 (1.69); 946	6.24 (2.04); 897	-0.13	(-0.28 to 0.02)	0.04
Cholesterol (mmol/l)	14	4.51 (1.05); 1388	4.67 (1.11); 1314	-0.15	(-0.20 to -0.10)	0.00
Low-density lipoprotein cholesterol (mmol/l)	14	2.69 (0.98); 1378	2.70 (0.98); 1299	-0.03	(-0.08 to 0.01)	0.00
High-density lipoprotein cholesterol (mmol/l)	14	1.15 (0.33); 1384	1.21 (0.39); 1312	-0.06	(-0.08 to -0.04)	0.00
Triglyceride (mmol/l)	14	1.73 (1.30); 1368	1.89 (1.51); 1297	-0.09	(-0.18 to -0.00)	0.01
Hb (g/l)	13	153.53 (14.71); 1291	143.58 (12.67); 1206	10.87	(8.19 to 13.55)	20.80
HbA1c (%)	8	6.46 (1.12); 748	6.58 (1.21); 742	-0.09	(-0.25 to 0.06)	0.03
HbA1c (%) sensitivity ^a	7	6.14 (0.94); 519	6.24 (1.08); 523	-0.89	(-2.43 to 0.64)	4.29
Haematocrit (%)	15	46.06 (4.37); 1399	42.94 (3.77); 1309	3.15	(2.42 to 3.88)	1.77
SBP (mmHg)	10	134.11 (17.14); 1069	133.31 (16.64); 1041	0.99	(-0.08 to 2.06)	0.00
DBP (mmHg)	10	77.20 (11.03); 1069	76.84 (10.98); 1041	0.48	(-0.30 to 1.26)	0.15

^a Sensitivity analysis excluding participants with diabetes.

Note

Outcomes were analysed using random effects.

precluded a meaningful evaluation of the impact of TRT on mortality and hampered the possibility of any subgroup analyses. Furthermore, the length of follow-up in most of the existing trials is likely to have precluded the accumulation of enough events. We were not able to assess whether the incidence of CV or CBV events was affected by the mode of administration of testosterone therapy as there were too few studies of oral testosterone. It is worth noting, however, that there have been no large studies directly comparing the effects of administration route on either TRT safety or efficacy in men with hypogonadism. There were inconsistent MACE classification and reporting within RCTs, and studies were of a relatively short duration to assess MACE risk (3–24 months). An ongoing trial, which is designed and powered to detect MACE during TRT for a period of 2 years in 6000 men with high risk of CV disease, will provide added clarity on the CV safety of TRT in MH.⁷⁶ IPD was retrieved from 62% of study groups eligible for inclusion. Aggregate meta-analysis suggested that outcome data were not significantly discrepant between studies with IPD and non-IPD. However, we cannot exclude that unreported MACE events in non-IPD studies would change the conclusions of our analysis.

We did not expect to find that neither age, nor BMI nor diabetes status was associated with the effectiveness of testosterone to improve symptoms compared with placebo in men with low testosterone. To explore these observations further, we conducted post hoc threshold analyses. These analyses revealed that the absolute levels of sexual function achieved during testosterone treatment were subject to thresholds related to age; men aged 50–70 and > 70 years had lower mean IIEF-15 compared with younger men. Furthermore, the absolute levels of sexual function achieved during testosterone treatment were lower in men with BMI above 30.6 kg/m² compared with leaner men. A

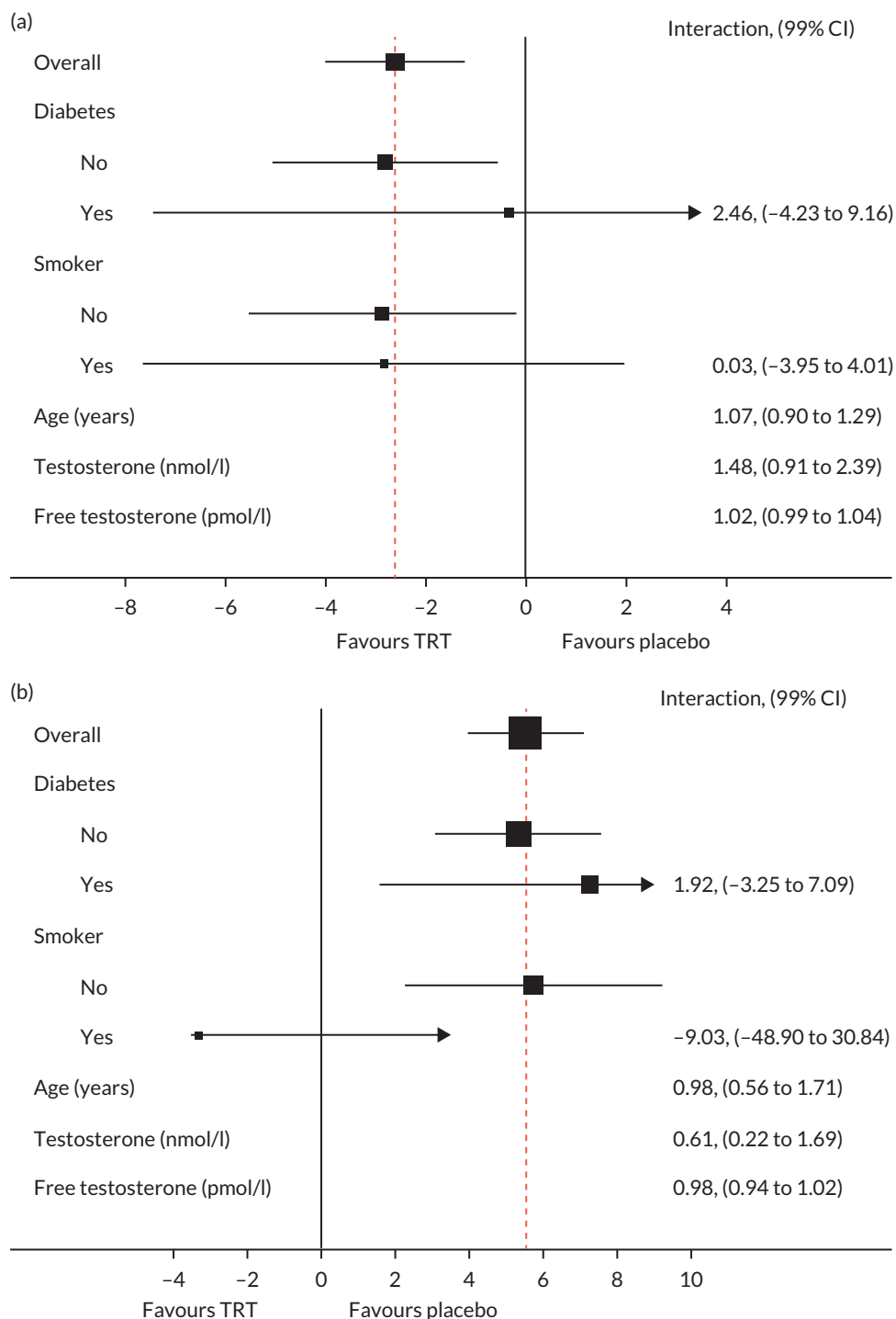


FIGURE 13 Subgroups for TRT vs. placebo. (a) AMS; (b) the IIEF-15; (c) the IIEF-5; (d) Hb (g/l); (e) BDI. (continued)

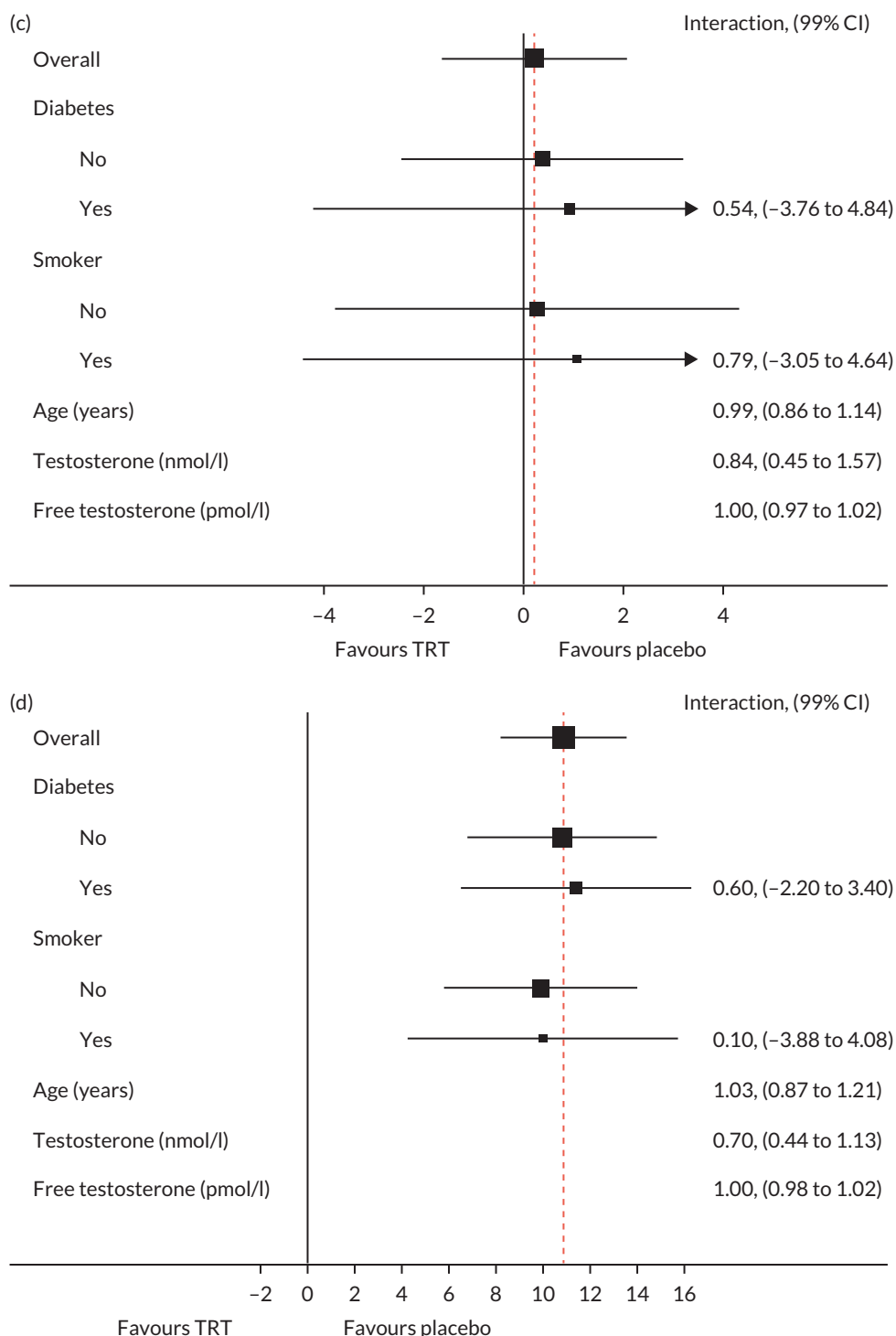


FIGURE 13 Subgroups for TRT vs. placebo. (a) AMS; (b) the IIEF-15; (c) the IIEF-5; (d) Hb (g/l); (e) BDI. (continued)

recent metaregression of aggregate data reported that increasing BMI was associated with reduced effect of testosterone compared on erectile function (one subscore of the IIEF-15 or IIEF-5²⁵). Our results may suggest a modification to this viewpoint: increments in sexual function during testosterone treatment are not significantly impaired by age or BMI. However, older men and men with obesity have poorer baseline sexual function which makes them less likely to have residual symptoms compared with other men.

There is no binary threshold of serum testosterone below which testosterone therapy is recommended. However, it may be logical to assume that men with more severely reduced serum testosterone would

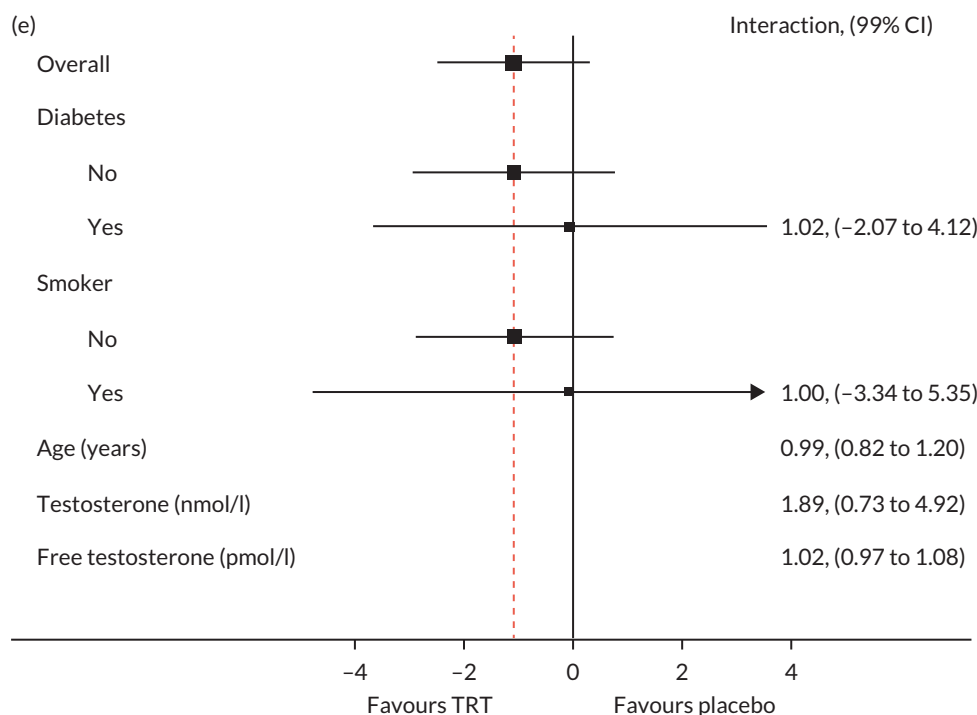


FIGURE 13 Subgroups for TRT vs. placebo. (a) AMS; (b) the IIEF-15; (c) the IIEF-5; (d) Hb (g/l); (e) BDI.

TABLE 11 Summary of other outcomes

Outcome	Number of studies	TRT n/N (%)	Placebo n/N (%)
Diabetes/diabetes complications	2	14/752 (1.9)	19/751 (2.5)
Prostate cancer	8	10/1293 (0.8)	3/1059 (0.3)
Oedema	7	34/1301 (2.6)	17/1290 (1.3)
Hypertension	7	28/1195 (2.3)	20/1182 (1.7)
High haematocrit	7	30/1079 (2.8)	5/993 (0.5)
Venous thromboembolism	4	5/1037 (0.5)	7/1034 (0.7)
Non-stroke CBV pathology ^a	3	4/655 (0.6)	11/648 (1.7)

a For example carotid occlusion and carotid stenosis.

have more severe symptoms which might be more amenable to correction by testosterone therapy. In keeping with this assumption, Corona *et al.*²⁵ reported that testosterone improved erectile function most effectively in RCTs with severely reduced baseline testosterone levels (< 8 nmol/l). We were, therefore, surprised to observe no significant interaction between serum testosterone and improved sexual function during testosterone treatment compared with placebo. We conducted post hoc threshold analysis to explore this further; the sexual function achieved during testosterone treatment was significantly higher in men with baseline serum testosterone > 9.8 nmol/l compared to other men. Taken collectively, our data suggest that increments in sexual function during testosterone treatment are not related to severity of low testosterone; hence, men with the most severe forms of low testosterone (and worse symptoms) are less likely to achieve symptom control compared with men with milder forms of low testosterone.

Conclusions

Results of this IPD meta-analysis have important implications for the management of patients with MH. We failed to find evidence that TRT increases risk of MACE in the short or medium term. In addition, we failed to identify subgroups at high risk of MACE during TRT. TRT effectively improves sexual symptoms and QoL outcomes in men with hypogonadism, even in older men, those with obesity and men with less severe forms of low testosterone. However, older men and men with obesity may anticipate more residual symptoms of low testosterone during testosterone treatment. These data provide some reassurance to patients with hypogonadism and their clinicians about the short- to medium-term safety of TRT use. Results of ongoing clinical trials will contribute to inform clinical practice on the long-term safety of TRT. Longer-term safety data on TRT will be available in the future from trials currently under way. However, our findings may help clarify the short- to medium-term safety of TRT for both clinicians and patients.

Chapter 3 Synthesis of qualitative and mixed-methods evidence evaluating men's experiences of low testosterone and/or health professionals' and care providers' views of testosterone replacement therapy

Introduction

Value of mixed-methods qualitative studies

Qualitative studies play a crucial role in the understanding of how patients experience aspects of, including symptoms related to conditions and interventions used to treat them. Studies using qualitative methods can explore the factors that facilitate or hamper the effectiveness of interventions, especially those that are patient-reported, and investigate how the delivery of interventions is perceived and implemented by users and providers.⁷⁷ The results of a quantitative review can be enriched and maximised by including a qualitative systematic review (i.e. by conducting a mixed-methods review). By combining qualitative and quantitative aspects, clinical and policy decision-makers can be better guided on the management of clinical conditions beyond issues of simple clinical effectiveness while taking into consideration perspectives and expectations of patients and providers and addressing potential barriers and challenges.

Role for qualitative studies in male hypogonadism

Approximately 30% of men aged 40–79 years have low levels of circulating testosterone. Besides its impact on sexual function and secondary sexual characteristics, including bone and muscle health, symptomatic low testosterone may also impair QoL, cognition, mental health and daily activities. RCTs have investigated the clinical effects of TRT in men with symptomatic low testosterone in heterogeneous patient groups (e.g. baseline testosterone, patient age, comorbidities) and using a range of validated symptom score questionnaires assessed impacts on QoL. A few qualitative studies have explored the perceptions of men who receive TRT, but information from these studies has not been systematically collected and summarised. This qualitative systematic review focuses on understanding men's experience of low testosterone, their views about the acceptability of TRT and/or the views of providers of their care. In addition to the overall aim of reviewing the evidence on men's, and their care providers', experiences of hypogonadism and its treatment, a further aim of this synthesis was then to compare the findings of what matters most to men to outcomes reported in existing comparative effectiveness trials (identified in our IPD meta-analysis, see [Chapter 2](#)) and to outcome reports in disease-specific PROMs (identified in [Chapter 4](#)).

Methods

Searching and identification of relevant studies

We developed a comprehensive search strategy, informed by relevant studies in the literature, to identify published papers reporting qualitative data from men with hypogonadism who received or considered receiving TRT, and papers reporting the views of care providers. An information scientist searched major electronic databases: Ovid MEDLINE, Embase, and PsycInfo, EBSCO CINAHL and Proquest ASSIA for papers published from 1992 (when TRT was introduced to practice) to February 2020. References of included studies were perused for further relevant papers. Search strategies are presented in [Appendix 1](#).

One review author (MA-M) independently screened all titles and abstracts, with a randomly selected sample of 10% cross-checked by a second review author (KG). A third author (JC) was consulted when consensus could not be reached regarding eligibility. We focused on primary studies that explored any aspect of TRT for testosterone deficiency from the perspective of men, their partners or their clinicians; this included living experience of testosterone deficiency, symptoms and treatment. Mixed-methods studies were included if the qualitative elements of the studies (i.e. methods and results) were reported separately.

We adopted the same eligibility criteria used for the IPD meta-analysis as part of the broader evidence synthesis. Specifically, we included studies that enrolled adult men (> 18 years old) with a diagnosis of hypogonadism confirmed by low levels of serum testosterone. We included studies that recruited participants seen in any clinical setting. Patients with hypogonadism due to genetic or congenital causes or to concurrent clinical conditions (e.g. Klinefelter syndrome, congenital hypogonadism, prostate cancer, etc.) were excluded from this qualitative review.

Data extraction and synthesis

Two review authors (MA-M and KG) independently read and extracted data from the included studies, shared notes and discussed study findings and interpretations during a series of meetings. Papers were initially organised alphabetically and subsequently grouped under emerging issues and themes. A data extraction form was developed and piloted for the purpose of this review. From each included study the following information was recorded: research question and context, objectives and methods, characteristics of participants, quotes from participants and interpretation of findings from study authors irrespective of whether it was supported by the participants' quotes.

Qualitative analysis

We conducted a thematic synthesis using both inductive and deductive approaches to analysis. According to current recommendations for thematic synthesis, we considered three stages to the

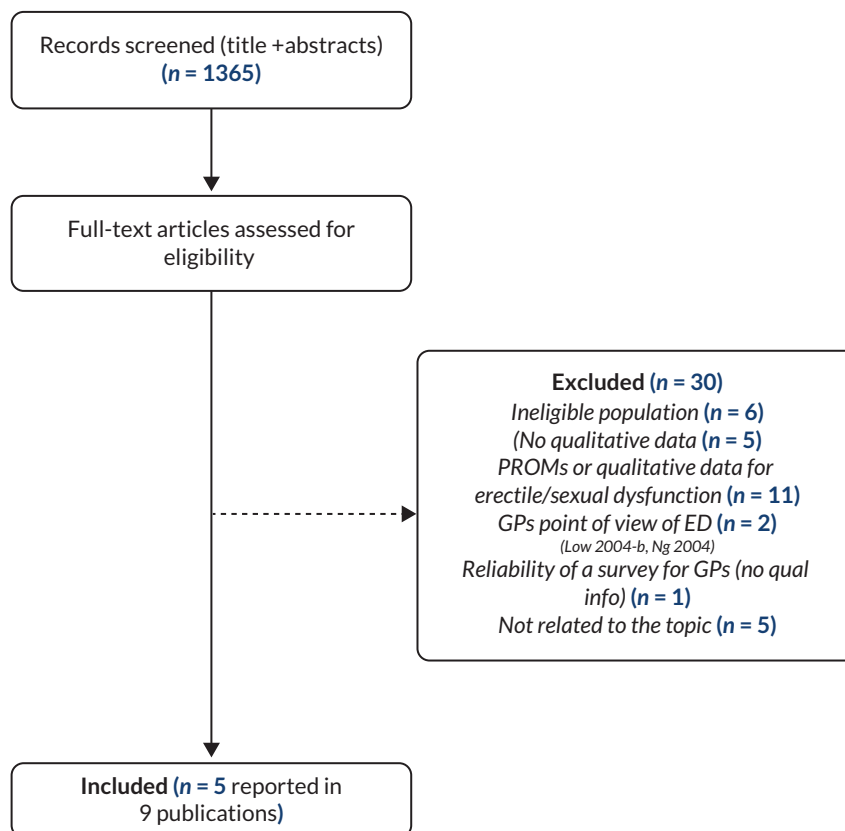


FIGURE 14 PRISMA flow diagram of qualitative study selection.

analysis.⁷⁸ First, we closely scrutinised the included studies to identify the main recurring themes and record the line-by-line coding of the qualitative findings; next, we organised the 'free codes' (i.e. single quotes) into related areas to construct 'descriptive' themes; and finally, we developed an 'analytical' theme. However, it was challenging to generate rich analytical themes beyond the original descriptive themes due to the lack of relevant data.

Quality assessment strategy

We appraised eligible studies for methodological rigour and theoretical relevance using the Critical Appraisal Skills Programme (CASP) tool.⁷⁹ Included studies were quality-appraised by one review author (MA-M), with a second author (KG) checking the completed assessments. We considered three main domains using 10 questions within the CASP tool: the validity of results, the results themselves (i.e. the findings) and the generalisability of results. Due to the small number of identified studies, we did not exclude any studies based on quality.

Assessment of confidence in the review findings

We applied grading of recommendations assessment, development and evaluation-confidence in the evidence from reviews of qualitative research (GRADE-CERQual) to the findings of the thematic synthesis.⁸⁰ The GRADE-CERQual approach is based on four components, which include the methodological limitations of included studies, the coherence of the review findings, the adequacy of data contributing to the review findings and the relevance of the included studies to the review question. Each finding of the thematic synthesis was assessed and discussed by two authors (MA-M and KG), who recorded any concern regarding any of the four GRADE-CERQual components before making an overall judgement of the confidence of the findings. We based our judgements on an initial assumption that all findings were 'high confidence' and were a reasonable representation of the phenomenon of interest, and then downgraded them accordingly if there were concerns regarding any of the GRADE-CERQual components.

Findings

Description of included studies

The literature search identified a total of 1365 citations. After screening the titles and abstracts of these citations, 39 studies were retrieved for full-text assessment. Thirty studies were subsequently excluded as they failed to meet our pre-specified inclusion criteria. Reasons for exclusion were ineligible populations ($n = 6$), focus on GPs' point of view or on a single symptom linked to hypogonadism such as erectile dysfunction ($n = 13$) or no relevant qualitative data reported ($n = 11$). Five studies, reported in nine publications, were included in this review (Figure 14).

The key characteristics of the five included studies are summarised in Table 12. Included studies were published between 2009 and 2016 and conducted in North America (USA = 4 and Canada = 1). None of the studies was linked to the trials included in the IPD meta-analysis (see Chapter 2). The five included studies reported data from patients with low testosterone who had or had not been treated with TRT. Only one study also reported data from healthcare providers' perspectives on the prescription of TRT.²⁸ Overall, the five studies provided data on the perspectives of 284 men (with the number of study participants ranging from 9 to 80) and 9 healthcare providers. All study participants were adult males; age ranged from 18 to 85 years. Mean age was not consistently reported across studies. The diagnostic criteria for hypogonadism were specified in three of the five included studies: two studies required a total serum TT level < 300 ng/dl (10.4 nmol/l) as entry criterion; in one study the majority of participants (22/26) had a total serum TT level < 300 ng/dl (10.4 nmol/l) while the remaining participants (4/26) had levels < 500 ng/dl. Depending on the aim of the study, some participants but not all were treated with TRT (see Table 12). The aim of three of the included studies was the development of PROMs; one study aimed to develop an instrument to identify men with hypogonadism and the other two studies explored influences on the rise in TRT.^{27,81,82}

TABLE 12 Participant characteristics of included studies

Study	Aim (as described within the papers)	Condition of focus	Participants' characteristics	Details of study	Qualitative methods
<p><i>First author:</i> Gelhorn <i>Year:</i> 2015 <i>Country:</i> USA⁸²</p>	<p>To develop a patient-reported outcome instrument, the Hypogonadism Impact of Symptoms Questionnaire (HIS-Q) and to assess its content validity. In a second publication (Gelhorn 2016), authors developed a briefer version of this same tool⁸³</p>	<p>Clinical diagnosis of hypogonadism (total serum TT level < 300 ng/dl) with or without TRT. The mean of the patients' lowest recorded testosterone levels was 184.9 ± 55.2 ng/dl, and the patients had been diagnosed with hypogonadism for 2.9 ± 3.9 years (range 0.3–20.6). Mean time since diagnosis (clinic report), years (SD) [range] 2.7 (2.6) [0.0–11.8]</p>	<p><i>Role:</i> patients <i>Number of participants interviewed:</i> 65 <i>Participant characteristics:</i> male participants, > 18 years old [mean 53.0 (SD 14.1)], with hypogonadism (mean serum TT level was 184.9 ± 55.2 ng/dl), able to read, speak and understand English. Participants with major health issues (e.g. endocrine, CV or mental disease) were excluded. <i>Socioeconomic and demographic characteristics:</i> 16.9% were Hispanic or Latino, 83.1% not Hispanic or Latino. Race reported as 1.5% American Indian or Alaska Native, 15.4% Black or African American, 75.4% white, 7.7% other; 86.2% were living with partner or spouse, family or friends</p>	<p><i>Recruitment:</i> participants were recruited through eight geographically diverse clinical sites in the USA. In Gelhorn 2016, participants were recruited through three clinical sites in New Jersey, New York and Washington State from November 2013 through November 2014. Unclear if the population overlaps between the two studies. <i>Further information:</i> The instrument development included a literature review, input from expert clinicians (<i>n</i> = 4) and qualitative study including the first phase with concept elicitation focus groups (5–8 participants each, <i>n</i> = 25); individual concept elicitation interviews by telephone (<i>n</i> = 5) or face-to-face (<i>n</i> = 9); and a subsequent phase including personal cognitive interviewing (<i>n</i> = 9) or electronic (<i>n</i> = 12). In Gelhorn 2016, a similar procedure was undertaken with fewer participants (<i>n</i> = 35)</p>	<p><i>Overall methods:</i> focus groups, one-on-one interviews. <i>Collection:</i> not reported for every phase. The four focus groups were conducted by the same experienced moderator (female) and trained assistant (female). <i>Framework or theory used:</i> not reported. <i>Analysis:</i> data from the interviews were analysed using thematic analysis. A saturation grid was developed to document the concepts endorsed by each participant or focus group</p>
<p><i>First author:</i> Hayes <i>Year:</i> 2015 <i>Country:</i> USA⁸¹</p>	<p>To establish the content validity of two new patient-reported outcome measures: Sexual Arousal, Interest, and Drive Scale and HED</p>	<p>Hypogonadism [either a prescription for low testosterone treatment or a laboratory sheet showing a TT level < 300 ng/dl (10.4 nmol/l)]</p>	<p><i>Role:</i> patients <i>Number of participants interviewed:</i> 72 <i>Participant characteristics:</i> men 18–85 years old with a diagnosis of hypogonadism</p>	<p><i>Recruitment:</i> participants recruited by a recruiting agency primarily through physician referrals and newspaper or internet advertisements between October 2010 and February 2012</p>	<p><i>Overall methods:</i> focus groups and individual in-depth interviews. <i>Collection:</i> the same interviewer (male) conducted all focus groups and the interviews. <i>Framework or theory used:</i> grounded theory</p>

TABLE 12 Participant characteristics of included studies (continued)

Study	Aim (as described within the papers)	Condition of focus	Participants' characteristics	Details of study	Qualitative methods
		No information reported on time since diagnosis	<i>Socioeconomic and demographic characteristics:</i> 90% were older than age 40 years, 63% whites and 93% had acquired hypogonadism as an adult; 40% had high blood pressure, 38% high cholesterol and 15% diabetes. Furthermore, 58% were receiving treatment (unclear if TRT)	<i>Further information:</i> they conducted four qualitative studies. Only study one was relevant to the current review and included concept elicitation (i.e. open-ended questioning to elicit concepts related to experiencing hypogonadism and its treatment). The interviews were scheduled to last 1 hour, and the focus groups 2 hours	<i>Analysis:</i> Broad topic area identification. Analysis conducted by two independent researchers
<i>First author:</i> Rosen <i>Year:</i> 2009 <i>Country:</i> USA ⁸⁴	To develop an instrument that could be used for identification of classification of men with hypogonadism	Hypogonadal patients (reported as clinical symptoms of hypogonadism as judged by a physician) and low level of TT. N = 26 controls N = 26 untreated hypogonadism N = 26 hypogonadism with TRT Of those with untreated hypogonadism: 22/26 had TT level < 300 mg/dl (10.4 nmol/l) 3/26 had testosterone level 300–400 mg/dl (10.4–13.9 nmol/l) 1/26 had testosterone level > 400 mg/dl (13.9 nmol/l) Months since diagnosis, treated patients = 50.4 (43.1), and untreated = 18.7 (23.3)	<i>Role:</i> patients <i>Number of participants interviewed:</i> 80 <i>Participant characteristics:</i> treated [receiving TRT; n = 26; testosterone mean 427 (SD 286) ng/dl] and untreated [no TRT in the past 3 months; n = 26; testosterone mean 258 (SD 75) ng/dl] diagnosed hypogonadal and eugonadal (control group, n = 28) patients from 21 to 74 years old, able to speak and read English, with cognitive competences and absence of any speech or comprehension difficulties. Patients with any major medical or psychiatric disorder were excluded. <i>Socioeconomic and demographic characteristics:</i> 83.7% were white, 10% were African American, 3.7% were Asian, and 2.5% were Native Hawaiian or other	<i>Recruitment:</i> recruited from different sources including physician providers, community-based services, health forums and media advertisements. Diagnosed hypogonadal patients (treated and untreated) were recruited from the practices of three physicians who are knowledgeable in the diagnosis and management of hypogonadism. Date not reported. <i>Further information:</i> they generated an item pool from focus groups (90–120 minutes) and in-depth interviews (45–90 minutes). Standardised scoring of the qualitative interviews was used to confirm conceptual domains to generate a questionnaire	<i>Overall methods:</i> data collection was through three focus groups (for each of the study groups), including 4–6 patients. Once the recruitment quota for each focus group was met, patients were invited for in-depth semistructured individual interviews. Inductive and deductive approaches and saturation approach were used. <i>Collection:</i> focus groups and interviews were led by a trained moderator (sex not reported). <i>Framework or theory used:</i> grounded theory. <i>Analysis:</i> Broad topic area identification. Analysis conducted by two researchers

continued

TABLE 12 Participant characteristics of included studies (continued)

Study	Aim (as described within the papers)	Condition of focus	Participants' characteristics	Details of study	Qualitative methods
<p>First author: Szeinbach Year: 2012 Country: USA²⁷</p>	<p>To create a final conceptual model and the Preference for the Testosterone Replacement Therapy (P-TRT) instrument</p>	<p>Participants who agreed to take part in research studies about TRT for conditions associated with a deficiency or absence of endogenous TT. All participants were recruited from a TRT manufacturer's mailing list since they were, or had been, taking TRT 'for conditions associated with a deficiency or absence of endogenous testosterone': that is, diagnosis of hypogonadism was not confirmed. In exchange for their participation, participants had the option to accept coupons toward their next purchase of a TRT product. Gives data on time on TRT – 299 days</p>	<p>Role: patients Number of participants interviewed: 58 Participant characteristics: male, aged > 18 years [mean 55 (SD 10) years], with current or previous experience using TRT, and be able to receive TRT at the time of the study. Socioeconomic and demographic characteristics: participants used TRT for an average of 175.0 ± 299.2 days. Four participants highlighted having problems with insurance coverage for TRT</p>	<p>Recruitment: participants were selected from a mailing list containing people who agreed to take part in research studies about TRT for conditions associated with hypogonadism. Enrolment via the online manufacturer-sponsored website was voluntary. Recruitment took place in December 2011. Further information: the instrument development included a literature review, input from expert clinicians and qualitative data. Firstly, a discussion guide was developed from the literature and expert opinion. Data were piloted, collected and coded from one-on-one from five participant interviews (lasting up to 1 hour). Then, one-on-one participant interviews (lasting up to 30 minutes) were conducted using the standard set of questions from the discussion guide. Afterwards, a group of experts (one physician, three researchers with extensive experience in psychometrics and a nurse practitioner with clinical experience with TRT) tested data and once consensus was reached on all possible items and themes, the final stage included the development of an instrument and conduct in-depth interviews</p>	<p>Overall methods: one-on-one participant interviews end expert's analysis to create an instrument to conduct in-depth interviews as part of the cognitive debriefing process. Researchers elicited and recorded responses from participants during interview sessions. Collection: interviewer(s) data not reported. Framework or theory used: grounded theory. Analysis: broad topic area identification. Transcription process included identification of recurring definitions and themes throughout the text, which produced rich descriptions and theoretical explanations of the concepts under investigation</p>

TABLE 12 Participant characteristics of included studies (continued)

Study	Aim (as described within the papers)	Condition of focus	Participants' characteristics	Details of study	Qualitative methods
<p>First author: Mascarenhas Year: 2016 Country: Canada²⁸</p>	To explore and describe factors that may influence the rise of prescribing and use of TRT on late-onset hypogonadism	<p>Patients: TRT users (67% had late-onset hypogonadism, the rest had different pathologies). Providers included primary care healthcare providers and specialists. Nine patients were recruited. All were on TRT. The diagnosis of hypogonadism was not confirmed. $N = 6$, late-onset hypogonadism; $n = 1$, HIV; $n = 1$ Klinefelter syndrome; $n = 1$ lymphoma. Years on TRT: $< 5 = 67%$; $5-15 = 22%$; and more than $15 = 11%$</p>	<p>Role: providers and patients Number of participants interviewed: 9 Number of providers interviewed: 13, from which six were primary care healthcare providers (three primary care physicians, two nurses and one pharmacist) and seven were specialist (five urologists and two endocrinologists). Participant characteristics: men > 18 years old; 45% of the participants were > 65 years old. Provider characteristics: all the professionals worked in an urban location, 91% were full-time health workers, and 47% had > 15 years in practice. Socioeconomic and demographic characteristics of participants: 55% were full-time employees, and the rest were unemployed</p>	<p>Recruitment: all participants (patients and providers) recruited from Ontario, through message distribution (fax, e-mail, social media) contacting clinician networks and circles of contact, posting flyers in clinics. Year not reported. Further information: each interview lasted from 30 to 60 minutes. Framework approach used and concepts identified from the literature were used to create a guide for the interviews</p>	<p>Overall methods: data identified from published literature and expert input. One-on-one semistructured telephone interviews. Collection: interviewer(s) data not reported. Framework or theory used: framework approach from Lewis 2003. Analysis: they developed a coding framework to include topics from raw data and previous concepts. Two analysts independently coded data</p>

Overall findings

From those studies that included patient-relevant data, six broad themes (with several linked subthemes) were identified in relation to men, and their care providers, experiences of low testosterone and receiving TRT as treatment (see [Table 13](#) for summary of overall findings and [Table 14](#) for how each paper contributed to the themes and subthemes). Five themes were ordered to reflect key timeline stages and decision points that a man with low testosterone may experience:

1. symptoms of low testosterone and impact on daily life
2. diagnosis of low testosterone
3. access to treatment information
4. perceived effects of TRT
5. expectations, experience and preference of the type of TRT.

The sixth theme on providers perceptions was reported separately. Each of these themes is presented below with findings presented according to the relevant participant group (i.e. patients, health professionals). The contribution of each paper to the themes and subthemes is presented in [Table 14](#).

TABLE 13 Summary of the key concepts identified in literature

Key concepts identified	Low testosterone symptoms and the impact such symptoms have in daily life	The diagnosis of low testosterone and access to treatment information	Access to treatment information	Perceived effects of TRT	Expectations, experience and preference of type of TRT
Overall description	In most of the studies, lack of energy, altered sleeping patterns, lack of strength, weight gain altered sexual activity/desire were the physical symptoms most reported from participants. Emotional/ affectional, cognitive and general well-being affects also reported. However, the frequency and severity of such symptoms poorly reported.	Two studies reported the perspective of patients regarding getting a diagnosis of HG, and the role and relevance of health professionals had in this process. However, this information was reported by the authors from the paper, rather than from quotes of participants. Szeinbach 2012 and Mascarenha 2016 reported that some participants understood the importance of testosterone monitoring and stated it would be easy to get this information from their physicians.	Some patients believe that their access to TRT information could facilitate their eventual use. The study in the USA by Szeinbach 2012 found that half of the participants described discovering TRT in different ways: either during a consultation with their general practitioner during a session for a related condition or through posters in their pharmacy and health professional practice, through friends and workers.	Most of the studies reported participants' perceptions of the effects of TRT on different symptoms, which mostly was positive perception towards the improvement of outcomes. However, some participants also reported no effect at all. Across studies, dosages, frequency and duration of TRT among participants were poorly or not described.	One study was designed to create a conceptual model and tool to test the preferences of participants with regard to the ease of TRT administration (Szeinbach 2012). Overall, participants preferred a product that was accessible to use, effortless and comfortable to apply, easy to handle, with accessible application location, and dried quickly.
Example(s)	<i>'... I woke up in the morning, I felt like I was more tired than when I went to bed ...' (Participant 01-103; Gelhorn 2015)</i> <i>'Loss of manliness' (no participant details; Rosen 2009)</i> <i>'I still look at women all the time, the beautiful ones. Mentally, it's like I still have it, but physically I don't have it like I used to ...' (age 54, adult onset; Hayes 2012)</i>	No quotes provided.	<i>'A couple [of] months ago, [I was] having some blood work done and read an article in Esquire magazine about TT. I asked my family doctor to have that checked'. No participant details; Mascarenha 2016</i>	<i>'feeling like myself again'. (No participant details; Szeinbach 2012)</i> <i>'My energy level's up; my libido's up'. (Participant 01-109; Gelhorn 2015)</i> <i>'[With the TT] I don't go soft. It I want to continue, I can continue'. (No participant details; Rosen 2009)</i> <i>'Not effective: I really was expecting like a boost of energy or some type of extra, sexual stamina/strength or something. I couldn't really feel much of anything'. (ID 10, 45 years old, average TRT use 113 days; Szeinbach 2012)</i>	<i>'Effective; pain to put it on every day; some burning sensations; wait time to dry'. [referring to injection TRT] (ID 2, 66 years old, average TRT use 120 days; Szeinbach 2012)</i> <i>'Effective: pleased with product; apply by myself; no transportation to doctor's office.' [referring to topical gel TRT]. (ID 1, 48 years old, average TRT use 90 days; Szeinbach 2012)</i> <i>'I don't use the gel any more. I didn't like having to wash my hands every time'. [referring to patch TRT]' (ID 9, 55 years old, average TRT use 365 days; Szeinbach 2012)</i>

TABLE 14 Contributions of included studies to themes and subthemes

Analytical themes	Symptoms of low testosterone and impacts on daily life									Perceived effects of TRT					Expectations, experience and preference of type of TRT						
	Altered sexual desire/activity	Lack of energy	Lack of strength	Altered sleeping patterns	Weight gain	Perceptions of masculinity	Cognitive function	Broader impacts on everyday life	Diagnosis of low TT	Access to treatment information	Sexual desire/activity outcomes	Strength/energy outcomes	Weight loss	Cognitive function outcomes	General well-being outcomes	Ease of Administration	Mode of administration	Beliefs about adverse effectiveness effects	Perceived costs	Providers' perceptions on TRT prescription	
Gelhorn 2016 (and 2016b)	● ^b	● ^c	● ^b	● ^b	● ^b	-	● ^b	● ^b	-	-	● ^c	● ^b	● ^a	● ^b	-	-	-	-	-	-	-
Hayes 2015	● ^c	● ^c	-	-	-	-	● ^b	● ^b	-	-	● ^c	-	-	-	-	-	● ^b	-	-	-	-
Rosen 2009	● ^a	● ^c	● ^c	● ^c	● ^c	● ^a	● ^a	● ^a	-	-	● ^a	● ^a	-	● ^a	● ^a	-	-	-	-	-	-
Szeinbach 2012	-	-	-	-	-	-	-	-	● ^b	-	-	● ^c	-	-	● ^a	● ^c	● ^a	● ^c	● ^c	● ^a	-
Mascarenha 2016	-	-	-	-	-	-	-	-	● ^b	● ^c	-	-	-	-	-	-	-	-	-	-	● ^c

- a Studies that provided quotes directly from participants that supported the theme/subtheme reported.
- b Studies that provided authors interpretations to support the theme/subtheme reported.
- c Studies that provided both quotes from participants and authors interpretations to support the theme/subtheme reported.

Note

No information provided by the publication.

Theme 1: Symptoms of low testosterone and impact on daily life

Within this concept, we considered any report of symptoms described by participants and perceived as associated with low testosterone. Several symptoms associated with low testosterone were reported across three of the included studies; these were broadly grouped into either physical or psychological domains.⁸¹⁻⁸⁴ For instance, the most frequently reported were physical symptoms (with five subthemes), which included a range of impacts from lack of energy, altered sleeping patterns, lack of strength, weight gain, altered sexual activity/desire and altered cognition (e.g. loss of memory).^{28,81,82,84}

Physical symptoms: altered sexual desire/activity

One of the most frequently cited subthemes of low testosterone symptomatology on physical function related to altered sexual desire and/or activity. In three of the included studies, participants described how low testosterone affected their sexual desire and activity and negatively impacted on their libido.^{81,82,84}

'... your sex drive is decreased. I'd have to tell them that your arousal level is greatly decreased ... I just wasn't excited like I have been in the past. And then even when [sexual] opportunities came ... I just wasn't still excited.'

(Age 52, adult-onset; Hayes 2015)

'I used to feel that I had an extremely active libido, and that went to a very low libido. So, I pretty much didn't initiate any kind of sexual activity. And then even my wife initiated it ...'

(No participant details; Rosen 2009)

With regard to desire, some participants also highlighted changes in sexual activity and/or performance, along with specific concerns around erectile dysfunction.

'I see stuff, like, I watch a porn video and I don't even get excited. I don't get erect or anything, and that's not like me ... nothing turns me on.'

(Age 48, adult onset; Hayes 2012)

'The hardness of an erection isn't the same; it dwindles.'

(No participant details; Rosen 2009).

The authors of one study reported that 'Spontaneously reported symptoms [of having low testosterone] included low sex drive or low sexual desire; inability to complete the sexual act; difficulty maintaining an erection; ... less intense climax or orgasm; and change in sensitivity of genitals' (Gelhorn 2016b).

The changes in desire and sexual activity highlighted the need the participants felt to satisfy their partners and the difficulties they had to face because of low testosterone, with some admitting they felt unable to 'do their part in that aspect of the relationship' (Rosen 2009). These broader perspectives of masculinity and their role in a relationship are also reported in a subsequent theme relating to perceptions of masculinity.

Physical symptoms: lack of energy

Participants described how a lack of energy negatively affected their activities throughout the day, with some reporting being exhausted even after a full night's sleep and others recognising it was worse in the evening. In general, a lack of energy was described to affect the ability to conduct 'normal' daily activities.

'... I was just tired. I just didn't have any energy ... I woke up in the morning, I felt like I was more tired than when I went to bed ... find yourself exhausted. And then on top of it now, I don't have that energy I used to.'

(Participant 01-103; Gelhorn 2016)

'Exhausted. For example, in the evening. Finish dinner and I say, oh, I'm exhausted. I go to do something, like screw a light bulb in, or I like to cook a little I stand at the sink a lot, [and] complain when I'm finished.'

(Study 1, participant age 81, adult onset; Hayes 2012)

'Completely exhausted. Could stay in the bed around the clock. Would even put off urinating as long as I could rather than get up and off the bed to go urinate, completely exhausted.'

(No participant details; Rosen 2009)

The authors of one of the included studies noticed that participants used different terms to describe their low energy levels such as 'lethargic', 'sluggish', 'physically drained'; however, the terms that resonated most often were 'tired' and 'totally exhausted' (Hayes 2012).

Physical symptoms: altered sleeping patterns

Two of the included studies reported that participants suffered from tiredness and sleep disturbances.⁸²⁻⁸⁴ In one study, participants complained of falling asleep during the day and reported having problems with night waking as well as difficulties going back to sleep.

'Just being tired all the time. It just doesn't seem right. Even coming off a vacation I felt tired and was off most of the holidays. A couple of days I ended up just falling asleep in the chair in the mid-afternoon.'

(No participant details; Rosen 2009)

'Typically, I don't have a hard time falling asleep. I have a hard time staying asleep, in the first hour or so. Typically, if I wake up within the first hour of falling asleep, I'm up for several hours. I can't get myself back to sleep.'

(No participant details; Rosen 2009)

The authors of one study reported that:

'The sleep disturbances participants (n = 36) described varied; they regularly woke up at night (n = 10; 28 %), had difficulty going back to sleep (n = 4; 11 %), or had poor quality sleep (n = 8; 22 %); nine of the men (25 %) reported increased napping.'

(Gelhorn 2016)

Physical symptoms: lack of strength

Some participants explained that one of the effects of low testosterone was a lack of physical strength, especially in relation to those activities they were able to carry out before.⁸²⁻⁸⁴ Participants also reported a reduction in their overall physical strength, which they attributed to low testosterone levels.

'Say you're carrying groceries and you pick them up, you can't hold them as long as you used to hold them The strength goes out of me quicker.'

(No participant details; Rosen 2009)

'Spontaneously reported symptoms [of having low TT] included ... inability to build muscle; lack of muscle strength or decreased muscle strength;'

(Authors' interpretations. Gelhorn 2016b)

Physical symptoms: weight gain

In contrast to concerns over lack of energy, tiredness and lack of strength with several quotes across studies, a less frequently raised concern was weight gain linked to low testosterone,⁸²⁻⁸⁴ with one participant stating:

'I kept insisting that my weight and my tenderness and everything else wasn't due to over-eating or over-drinking or lack of exercise ... I was working out 4 days a week. I was running five miles. I was playing squash 7 days a week ... I was getting heavier and heavier.'

(No participant details; Rosen 2009).

The authors of another study indicated that 'Spontaneously reported symptoms [of having low testosterone] included ... trouble losing weight; gaining weight more easily ...' (Gelhorn 2016b).

Across included studies, participants also described some psychological effects of low testosterone, which were grouped into three subthemes: perception of masculinity,⁸⁴ cognitive function⁸¹⁻⁸⁴ and broader impact on everyday life.⁸¹⁻⁸⁴

Psychological symptoms: perceptions of masculinity

In one of the included studies, participants reflected on their sense of masculinity, explaining they felt a sense of a 'loss of manliness' or 'less of a man' which was implicitly associated with their sexual performance/function.⁸⁴

'Being a man is just being a man. Just, you know. Being alive ... Being a man in the sense of ... having a good time, keeping your partner happy. Just enjoying life. And that's one part that being a man that I'm not enjoying.'

(No participant details; Rosen 2009)

Psychological symptoms: cognitive function

Among the psychological consequences of low testosterone, the subtheme related to cognitive function includes perceived changes in memory, concentration or attention. In one study, participants described the effects of low testosterone on their memory, complaining about their inability to maintain the thread of a story when reading a book.

'I used to ... read a book in 2 days and tell you everything about it. Can't do that anymore. I don't really want to read a book any more, because I have to keep going back over and over.'

(No participant details; Rosen 2009)

The authors of another study reported that when considering impacts on cognition participants reported problems with the following:

'motivation (n = 16; 44%), loss of interest (n = 11; 31%), problems with memory/forgetfulness (n = 11; 31%), problems with focus/concentration (n = 6; 17%), less drive/ambition (n = 3; 8%), short attention span (n = 3; 8%), and indecisiveness (n = 1; 3%).'

(Gelhorn 2016)

Psychological symptoms: broader impacts on everyday life and general well-being

Men also reported the impacts of low testosterone on other aspects of their lives, which affected their confidence, lack of motivation and generally feeling low.

'... lost a position at work ... lost my motivation to succeed. I lost my energy to go the extra mile to get projects done and stand up for what I believed in Performance reviews and stuff were low for the very first time in my career.'

(No participant details; Rosen 2009)

The authors of one of the included studies reported that

'Many of the men (n = 36) reported having less confidence or lower self-esteem (n = 10; 28%) Few men reported symptoms such as feeling mellow, introversion, feeling alone, fear of rejection, anxiety and being moody, emotional or sensitive.'

(Gelhorn 2016)

Theme 2: Diagnosis of low testosterone

Any account of the challenges experienced by the participants with regard to the diagnostic strategies used to establish the presence of hypogonadism was recorded in this theme. Two studies reported the participants experience of getting a diagnosis of hypogonadism and the role of health professionals in this process.^{27,28} It is worth noting that this information was obtained from the authors' reporting of participants' experience, rather than from quotes obtained directly from the participants. Szeinbach 2012 and Mascarenha 2016 reported that participants understood the importance of testosterone monitoring and stated it would be easy to obtain this information from their physicians.^{27,28} Mascarenha 2016 also discussed the persistence of some participants, defined as 'drug seekers', to acquire and use TRT, irrespective of the advice of their physicians.²⁸ These 'persistent' patients were reported to have consulted multiple physicians until they found one ready to prescribe them TRT (regardless of the diagnosis). Mascarenha 2016 reported also that one participant who did not feel satisfied with his physician's advice chose to increase his dose of TRT and subsequently, when he failed to perceive any immediate effects, requested switching products.²⁸

'Both patients and providers participants mentioned that they know of primary care physicians or specialists who prescribe TRT without testing for low testosterone levels and based on informal discussions or e-mail communication.'

(Authors' interpretation, Mascarenha 2016)

'While only two participants were able to recall their testosterone levels, the other three participants understood the importance of testosterone monitoring and stated it would be easy to obtain this information from their physicians.'

(Authors' interpretation, Szeinbach 2012)

Theme 3: Access to treatment information

In this theme, we recorded any account of participants' experience of access to information on TRT. Some participants explained that access to TRT information could facilitate their eventual use of the therapy. Szeinbach 2012 observed that participants came to receive TRT via different routes: during a consultation [e.g. with their general practitioner (GP) regarding a related condition]; through posters at their pharmacy; through friends and co-workers; popular magazines; internet searching.²⁷

'A couple [of] months ago, [I was] having some blood work done and read an article in Esquire magazine about testosterone. I asked my family doctor to have that checked.'

(No participant details; Mascarenha 2016)

Similarly, Mascarenha *et al.* reported that some participants felt that the marketing and advertisements 'spoke to' their perceived needs.²⁸ In particular, participants considered the information on improved sexual function and energy levels of particular interest. Mascarenha *et al.* maintained also that 'While most patient participants found it easy to access information on the positive effects of TRT and how to acquire it, they seem to have little knowledge about its side effects or risks'. Some participants also conveyed the desire to receive more information from health professionals on the availability and effects of TRT.²⁸

Theme 4: Perceived effects of TRT

In this theme, we considered changes in symptoms that participants perceived as associated with TRT. Several perceived effects attributed to TRT were reported across four of the included studies; as per theme 1, these were broadly grouped into either physical or psychological symptom-modification.^{27,81-84}

We identified five interlinked subthemes, specifically; sexual desire/activity, strength/energy, emotional or affectional or well-being, cognitive function and general well-being. The most reported perceived effects were physical symptoms (three subthemes), which included a range of impacts including sexual activity outcomes, strength and energy improvement.

It is important to acknowledge that information on dosages, frequency and duration of TRT among participants was poorly reported or not reported across included studies, except for the study by Szeinbach 2012, where the TRT type, dosage and time of use were specified for each patient. For this reason, some perceptions regarding testosterone therapy and its effects might differ and might also influence the response of participants based on the dosing and duration of treatment.

Physical symptoms: sexual desire/activity

One of the most frequently cited subthemes was the perceived effects of TRT on sexual desire/activity. Three of the included studies reported that participants described how TRT improved their sexual desire and activity and impacted on their libido and sexual performance.

'I have more desire than I did for a long time.'

(Participant 01-108; Gelhorn 2016)

'My energy level's up; my libido's up.'

(Participant 01-109; Gelhorn 2016)

'... the erections were better, sex was better, ejaculations were better; I started noticing a good difference, high energy; I was keeping the weight down.'

(Participant 02-104; Gelhorn 2016)

'If I have the (testosterone) shot, there's no reduction in the desire. If you don't have the shot, then you have no desire.'

(No participant details; Rosen 2009)

Hayes 2012 reported that 'Several of the participants who were currently receiving treatments such as solution applications noted an increase in their sex drive'.

On the other hand, some participants did not experience positive changes in sexual desire and performance after using TRT.

'Not effective: I really was expecting like a boost of energy or some type of extra, sexual stamina/strength or something. I couldn't really feel much of anything.'

(ID 10, 45 years old, average TRT use 113 days; Szeinbach 2012)

Physical symptoms: strength/energy

In three studies, participants explained their need for an 'energy boost' to alleviate their low testosterone symptoms and described their perceived improvement in energy levels after TRT.²⁷ Some participants reported 'feeling more muscular' and commented on improved muscle strength and energy levels throughout the day.⁸²

'Very good. It gives you the energy you need.'

(ID 16, 62 years old, average TRT use 1460 days; Szeinbach 2012)

'... The shots [of TRT] really hype you up, puts you almost on a cocaine buzz.'

(ID 8, 47 years old, average TRT use 120 days; Szeinbach 2012)

'... I started noticing a good difference, high energy; I was keeping the weight down.'

(Participant 02-104; Gelhorn 2016)

The authors of one study pointed out that:

'The majority of the participants noticed changes in their energy level and an increased libido after starting testosterone replacement therapy.'

(Gelhorn 2016)

Nevertheless, there were some participants who did not experience the anticipated positive effects of TRT on energy level.

'Not effective: I really was expecting like a boost of energy or some type of extra, sexual stamina/strength or something. I couldn't really feel much of anything.'

(ID 10, 45 years old, average TRT use 113 days; Szeinbach 2012)

Physical symptoms: weight loss

Gelhorn 2016 reported that one participant noted some weight loss as a positive effect of TRT probably linked to the improved energy levels.

'... the erections were better, sex was better, ejaculations were better; I started noticing a good difference, high energy; I was keeping the weight down.'

(Participant 02-104; Gelhorn 2016)

Psychological symptoms: general well-being

Three of the included studies reported impacts of TRT on general well-being. Szeinbach *et al.* reported that some of the participants they interviewed experienced changes in general well-being often expressed as 'feel like myself again'. One participant observed an improvement in self-esteem as a result of being more energetic and masculine.⁸⁴

'... one of the biggest benefits [TRT] I get is self-esteem, because there's more energy and feeling more muscular and masculine. And that goes away when I'm not on the testosterone ...'

(No participant details; Rosen 2009)

Another participant noted that not all problems improved after TRT and acknowledged that some symptoms can be interrelated.

'Helped as far as my energy level. I don't know if it has helped with regard to erectile dysfunction, I don't know which part was mental and physical.'

(ID 7, 54 years old, average TRT use 365 days; Szeinbach 2012)

One participant reported a broader range of symptoms and recognised the relatedness and interplay that may exist between them.⁸⁴ Some of these symptoms included psychological (e.g. anxiety), emotional (e.g. self-esteem), or well-being (e.g. masculinity perceptions) outcomes that were reported as improved after the therapy.

'I attribute a lot of the depression and anxiety to lack of self-esteem, which comes with testosterone ... one of the biggest benefits [TRT] I get is self-esteem, because there's more energy and feeling more muscular and masculine. And that goes away when I'm not on the testosterone.'

(No participant details; Rosen 2009)

Theme 5: Expectations, experience and preference of type of TRT

Three studies reported participants' expectations, experience and preferences with regard to the type of TRT. Five subthemes were identified across the included studies, including: ease of administration, mode of administration, beliefs about effectiveness, perceived adverse effects, and costs.

Ease of administration

One of the included studies was designed to create a conceptual model and tool to test the preferences of participants for ease of administration of TRT.²⁷ This study assessed the experiences and perceptions of participants for different types of TRT (i.e. gel vs. injections vs. patches). Participants used TRT for an average of 175.0 ± 299.2 days. Overall, participants expressed preferences for a product that was 'accessible to use', 'effortless' and 'comfortable to apply', 'easy to handle', 'with accessible application location' and 'that dried quickly'.

'The first theme, ease of use, encompassed all topical characteristics associated with testosterone gel products. Participants preferred a product that was convenient to use, easy to apply, easy to handle, with accessible application location, and dried quickly.'

(Authors interpretations, Szeinbach 2012)

Mode of administration

Participants from two of the included studies reported varied perspectives about the mode of administration of the TRT; preferences were highlighted for crucial features of the route of delivery which were linked back to ease of administration and perceptions about effectiveness.^{27,81}

'I used another product where I had to do the injection into the muscle, and the gel is easier because there is no sticking and blood, etc. But the injection more potent; lasts longer,'

(ID 4, 54 years old, average TRT use 365 days; Szeinbach 2012)

'I don't use the gel any more. I didn't like having to wash my hands every time.' [referring to patch TRT]

(ID 9, 55 years old, average TRT use 365 days; Szeinbach 2012)

'Overall, I guess it would be a fair experience. Well, as opposed to injections and other products I've used, I guess the gel's downfall is that you had to wait for it to dry. It wasn't a noticeable boost; the boost was more gradual.' [referring to topical gel TRT]

(ID 14, 42 years old, average TRT use 90 days; Szeinbach 2012)

'For those who were receiving testosterone injections, they observed a dramatic increase at the beginning of the injection period, with a waning in their drive as time drew nearer for their next injection.'

(Authors interpretations, Hayes 2012)

Beliefs about effectiveness

Only one of the included studies reported participants' beliefs about effectiveness of different types of TRT. Mixed views on effectiveness were reported.

'... pleased with product; apply by myself; no transportation to doctor's office' [referring to topical gel TRT].

(ID 1, 48 years old, average TRT use 90 days; Szeinbach 2012)

'... Mixed - the gel works and sometimes it doesn't. My testosterone level has fluctuated, I had had better results with injecting myself, but it is a painful and longer process. Patch leaves giant red marks; topical gel was less robust than injection.'

(ID 17, 48 years old, average TRT use 1825 days; Szeinbach 2012)

Perceived adverse effects

Two studies reported that participants expressed concerns about adverse effects associated with the TRT. One of these studies reported specific reactions such as rash or itching, or pain following application of the product (i.e. topical gel).

'I didn't like it at all. I was rather annoyed with working with it. Well I didn't like the time that it take to dry. And then I was running into rash and problems with itching. Never saw results with topical gel.' [referring to topical gel TRT]

(ID 12, 66 years old, average TRT use 90 days; Szeinbach 2012)

'Overall, it's decent, it irritates the skin but other than that it works well.' [referring to topical gel TRT]

(ID 19, 45 years old, average TRT use 180 days; Szeinbach 2012)

Costs

Szeinbach 2012 reported that the cost of treatment was among the variables participants took into consideration when expressing their preferences for products.²⁷ Some participants, for example, described how features of their insurance plans (e.g. copay help programmes to top up the cost of the preferred treatment) influenced the type of treatment they 'preferred' by default.

'First I found it very expensive; my insurance didn't cover it at all. I did find that it worked fine. I almost liked it better than the shot; it gave me a normal feel. The shots really hype you up, puts you almost on a cocaine buzz.' [referring to injection TRT].

(ID 8, 47 years old, average TRT use 120 days; Szeinbach 2012)

Szeinbach 2012 also noted the interplay between cost and mode of administration, stating 'one of these participants preferred an injection every 2 weeks compared with a product that required daily application, while another participant based his preference on product cost'.²⁷

Theme 6: Providers' perceptions on TRT prescription.

The final theme relates to the perspectives of health professionals who prescribe TRT. Only one of the included studies, Mascarenha *et al.*, conducted in Canada, assessed the perspectives of health providers (all types of clinicians, i.e. primary care physicians, nurses, pharmacists) about prescription of TRT.²⁸ In this study, providers expressed a high level of uncertainty and 'diagnostic ambiguity' with regard to the diagnosis of hypogonadism – particularly of late-onset hypogonadism (LOH) – and subsequent prescription of TRT, mainly due to the non-specificity of its symptoms. Some providers explained that there is little consensus on what constitutes low and normal levels of testosterone, described the diagnosis of asymptomatic patients as challenging, and suggested that cut-offs for normal ranges may vary between individuals.

'Is your current testosterone too low for you? Or is it too low for what you are used to?'

(Primary care physician, Mascarenhas 2016).

It is worth noting that this study focused exclusively on the perspectives of health providers who prescribed TRT in their daily routine and did not include the influences of a non-prescription approach. In general, providers indicated that 'clinical guidelines on the interpretations and administration of tests were perceived to be vague' and described different preferences in terms of tests used to determine the level of testosterone (e.g. total serum testosterone levels versus bioavailable testosterone levels). Non-specialist providers (i.e. primary care clinicians) did not consider the timing of the test or repeat testing as critical for the diagnosis. Some providers were more inclined to request a total serum testosterone test because it was covered by the participant's personal insurance plan and expressed some scepticism about the accuracy of bioavailable testosterone tests conducted in private laboratories. Some providers recognised that the number of patients who were keen to try TRT was increasing and reported also being aware of colleagues who prescribed TRT without testing the participants' testosterone level.

While some providers expressed the desire to exclude other clinical conditions before starting to prescribe TRT, others preferred to start TRT straight away and monitor its effects before deciding whether testing for other clinical conditions was necessary.

In general, providers noted a rise in the availability of information on TRT, particularly the use of the concept 'andropause', exploited for marketing reasons. Some providers believed that there was not enough evidence to compare low levels of testosterone in men with the concept of 'menopause' in women.

Providers also described how the decision about the appropriateness of TRT was influenced by the evidence (or lack of evidence) on its safety. Some providers claimed that there were 'myths' about TRT safety, particularly about the potential harms or side effects of the use of TRT in patients with certain clinical conditions (i.e. prostate cancer). Others preferred a more cautious approach due to the lack of evidence on the long-term consequences of TRT.

'I can see how someone might see the latest studies and say "my God, this is proof that [TRT] are dangerous". Someone like me, who follows the literature closer, understands the potential risks and potential benefits.'

(Endocrinologist, no further detail provided)

Across providers, there were different perspectives on the 'appropriateness of TRT'. While some specialists (e.g. oncologists) expressed the view that TRT had to be reserved only for 'profoundly low' cases, some primary care clinicians and specialists with an interest in men's health believed that appropriateness of TRT could vary according to the symptoms and general health of patients and their testosterone serum results. Other primary care physicians and general urologists maintained that TRT was appropriate for treating any symptomatic patient with a low testosterone test result, regardless of their underlying clinical conditions.

Quality assessment results

The methodological quality of the five included studies was assessed using the CASP tool. As the included studies sought to interpret or illuminate the actions and/or subjective experiences of the recruited participants, their findings were considered valid and relevant to address the research question of this qualitative synthesis.

The research design varied across included studies. With the exception of the study by Mascarenha *et al.*, in which the authors did not justify or discuss the choice of the study design, all remaining studies provided a justification and rationale for the choice of their study design.²⁸ Apart from the study by Gelhorn *et al.*, the recruitment strategy and setting were explained in all remaining studies. Three studies provided information on the relationship between the researchers and the participants.^{27,81,84} In the remaining two studies, the researchers did not critically assess their role and potential influence during the study.^{28,82}

The study by Gelhorn *et al.* was considered at potential risk of bias as it was sponsored by a pharmaceutical company who provided study support to some of the authors.⁸² As the role of the funder in the data analysis and drawing of conclusions was not clarified in the study, it is unclear whether the findings were interpreted in an objective and independent way.

All the studies discussed the contribution of their findings to existing knowledge and understanding. Overall, the findings across the five included studies were valuable and of acceptable quality and given the small number of studies we chose not to exclude.

Confidence in the findings

Our confidence in the findings of this qualitative evidence synthesis was assessed using the GRADE-CERQual approach (see [Table 15](#)). Findings were downgraded for 'methodological limitations' due to the lack of reported researcher reflexivity (i.e. how their own personal experiences may influence interpretations) across all studies, which may be particularly important when researchers receive support from the pharmaceutical industry. We typically downgraded a finding for concerns about 'coherence'

TABLE 15 CERQual evidence profile

Summary of review finding		Studies contributing to review finding	Methodological limitations	Coherence	Adequacy	Relevance	CERQual assessment of confidence in the evidence
Theme 1: Symptoms of low testosterone and impacts on daily life							
1	Altered sexual desire/activity	Gelhorn 2016 (and 2016b) Hayes 2015 Rosen 2009	Moderate concerns about methodological limitations, one study did not adequately address the recruitment strategy or analysis.	No concerns about coherence.	Minor concerns about adequacy. Three studies offered moderately rich data. Data retrieved come from direct participants quotes and some from authors' interpretation.	Moderate concerns about relevance given that most included population were white.	Moderate confidence
2	Lack of energy	Gelhorn 2016 (and 2016b) Hayes 2015 Rosen 2009	Moderate concerns about methodological limitations, one study did not adequately address the recruitment strategy or analysis.	No concerns about coherence.	Minor concerns about adequacy. Three studies offered moderately rich data. Data retrieved come from direct participant quotes and some from authors' interpretation.	Moderate concerns about relevance given that most included population were white.	Moderate confidence
3	Lack of strength	Gelhorn 2016 (and 2016b) Rosen 2009	Moderate concerns about methodological limitations.	Minor concerns about coherence. Some data slightly ambiguous.	Moderate concerns about adequacy. Two studies offered relatively limited data. Data retrieved come from direct participant quotes and some from authors' interpretation.	Moderate concerns about relevance given that most included population were white.	Moderate confidence
4	Altered sleeping patterns	Gelhorn 2016 (and 2016b) Rosen 2009	Moderate concerns about methodological limitations.	Minor concerns about coherence. Some data slightly ambiguous.	Moderate concerns about adequacy. Two studies offered relatively limited data. Data retrieved come from direct participant quotes and some from authors' interpretation.	Moderate concerns about relevance given that most included population were white.	Moderate confidence
5	Weight gain	Gelhorn 2016 (and 2016b) Rosen 2009	Moderate concerns about methodological limitations.	Minor concerns about coherence. Some data slightly ambiguous.	Moderate concerns about adequacy. Two studies offered relatively limited data. Data retrieved come from direct participant quotes and some from authors' interpretation.	Moderate concerns about relevance given that most included population were white.	Moderate confidence
							continued

TABLE 15 CERQual evidence profile (continued)

Summary of review finding	Studies contributing to review finding	Methodological limitations	Coherence	Adequacy	Relevance	CERQual assessment of confidence in the evidence	
6	Perceptions of masculinity	Rosen 2009	No concerns about methodological limitations.	No concerns about coherence.	Moderate concerns about adequacy because of relatively limited data.	Moderate concerns about relevance given that most included population were white.	Moderate confidence
7	Cognitive function	Gelhorn 2016 (and 2016b) Hayes 2015 Rosen 2009	Moderate concerns about methodological limitations, one of the studies did not adequately address the recruitment strategy or analysis.	No concerns about coherence.	Moderate concerns about adequacy. Two studies offered relatively limited data. Data retrieved come from direct participant quotes and some from authors' interpretation.	Moderate concerns about relevance given that most included population were white.	Moderate confidence
8	Broader affects on everyday life	Gelhorn 2016 (and 2016b) Hayes 2015 Rosen 2009	Moderate concerns about methodological limitations, one study did not adequately address the recruitment strategy or analysis.	No concerns about coherence.	Moderate concerns about adequacy. Two studies offered relatively limited data. Data retrieved come from direct participant quotes and some from authors' interpretation.	Moderate concerns about relevance given that most included population were white.	Moderate confidence
Theme 2: Diagnosis of low TT							
9	Diagnosis of low TT	Szeinbach 2012 Mascarenha 2016	Moderate concerns about methodological limitations, one study was overall poor quality.	Minor concerns about coherence. Some data slightly ambiguous.	Moderate concerns about adequacy. Two studies offered relatively limited data. Data retrieved come from authors' interpretation.	Significant concerns about relevance. Neither study reported ethnicity.	Low confidence
Theme 3: Access to treatment information							
10	Access to treatment information	Mascarenha 2016	Significant concerns about methodological limitations, included study was overall poor quality.	No concerns about coherence.	Moderate concerns about adequacy. Offered relatively limited data with the majority of data from authors' interpretation.	Significant concerns about relevance. Study did not report ethnicity.	Low confidence

Summary of review finding		Studies contributing to review finding	Methodological limitations	Coherence	Adequacy	Relevance	CERQual assessment of confidence in the evidence
Theme 4: Perceived effects of TRT							
11	Sexual desire/ activity outcomes	Gelhorn 2016 (and 2016b) Hayes 2015 Rosen 2009	Moderate concerns about methodological limitations, one of the studies did not adequately address the recruitment strategy or analysis (reflexivity was not addressed in the two studies, which may be particularly important as funded by pharmaceutical industry).	No concerns about coherence.	Minor concerns about adequacy. Three studies offered moderately rich data. Data retrieved come from direct participant quotes and some from authors' interpretation.	Moderate concerns about relevance given that most included population were white.	Moderate confidence
12	Strength/ energy outcomes	Gelhorn 2016 (and 2016b) Rosen 2009 Szeinbach 2012	Moderate concerns about methodological limitations, one study did not adequately address the recruitment strategy or analysis (reflexivity was not addressed in one study, which may be particularly important as funded by pharmaceutical industry).	No concerns about coherence.	Minor concerns about adequacy. Three studies offered moderately rich data. Data retrieved come from direct participant quotes and some from authors' interpretation.	Moderate concerns about relevance given that most included population were white and one study did not report ethnicity.	Moderate confidence
13	Weight loss	Gelhorn 2016 (and 2016b)	Moderate concerns about methodological limitations did not adequately address the recruitment strategy or analysis.	No concerns about coherence.	Moderate concerns about adequacy because of limited data.	Moderate concerns about relevance given that the majority of included population were white.	Moderate confidence
14	Emotional/ affectional/ well-being outcomes	Rosen 2009	No concerns about methodological limitations.	No concerns about coherence.	Moderate concerns about adequacy because of limited data.	Moderate concerns about relevance given that most included population were white.	Moderate confidence
15	Cognitive function outcomes	Gelhorn 2016 (and 2016b) Rosen 2009	Moderate concerns about methodological limitations.	Minor concerns about coherence. Some data slightly ambiguous.	Moderate concerns about adequacy because of limited data.	Moderate concerns about relevance given that most included population were White.	Moderate confidence
16	General well-being outcomes	Szeinbach 2012	Minor concerns about methodological limitations.	No concerns about coherence.	Moderate concerns about adequacy because of limited data.	Moderate concerns about relevance. Study did not reported ethnicity.	Moderate confidence
							continued

TABLE 15 CERQual evidence profile (continued)

	Summary of review finding	Studies contributing to review finding	Methodological limitations	Coherence	Adequacy	Relevance	CERQual assessment of confidence in the evidence
17	Ease of administration	Szeinbach 2012	Minor concerns about methodological limitations.	Minor concerns about coherence. Some data slightly contradictory.	Moderate concerns about adequacy because of limited data.	Moderate concerns about relevance. Study did not report ethnicity.	Moderate confidence
18	Perceived adverse effects	Szeinbach 2012 Mascarenha 2016	Moderate concerns about methodological limitations, one study was overall poor quality.	Minor concerns about coherence. Some data slightly contradictory.	Moderate concerns about adequacy. Two studies offered relatively limited data. Data retrieved come from authors' interpretation.	Significant concerns about relevance. Neither study reported ethnicity.	Low confidence
19	Beliefs about effectiveness	Szeinbach 2012	Minor concerns about methodological limitations.	Minor concerns about coherence. Some data slightly contradictory.	Moderate concerns about adequacy because of limited data.	Moderate concerns about relevance. Study did not report ethnicity.	Moderate confidence
20	Mode of administration	Hayes 2015 Szeinbach 2012	Moderate concerns about methodological limitations.	Minor concerns about coherence. Some data contradictory.	Minor concerns about adequacy. One study offered relatively limited data. Data retrieved come from direct participant quotes and some from authors' interpretation.	Moderate concerns about relevance. Only one study reported ethnicity, and most of the participants were white.	Moderate confidence
21	Costs	Szeinbach 2012	Minor concerns about methodological limitations.	No concerns about coherence.	Moderate concerns about adequacy because of limited data.	Moderate concerns about relevance. Study did not report ethnicity.	Moderate confidence
22	Providers perceptions on TRT prescription	Mascarenha 2016	Significant concerns about methodological limitations included study was overall poor quality.	Minor concerns about coherence. Some data contradictory.	Moderate concerns about adequacy. Offered relatively limited data with most data from authors' interpretation.	Significant concerns about relevance. Study did not report ethnicity.	Low confidence

when there were some concerns about discrepancies between the data from primary studies. Downgrading due to data 'adequacy' occurred when we had concerns about the richness or quantity of the data from included studies supporting a review finding. Most of the studies were also downgraded for 'relevance', because sociodemographic characteristics of included participants were poorly reported across included studies.

Discussion

This evidence synthesis has combined and summarised findings from qualitative studies in the literature that collected and reported data from men with low testosterone (i.e. having hypogonadism) and their care providers. The resulting synthesis of data from five included studies identifies a range of important considerations for men at key decision points across the timeline of living with low testosterone: starting with diagnosis and ending with experiences of treatment. Within the broad themes that map to these decision points, we identified several interconnected subthemes highlighting the complexity with regard to how symptoms influence many aspects of men's lives and their experiences of treatment. This process is likely not linear, with some participants circling back to seek additional information if perceived effectiveness of one type of TRT is not met, and some participants might not experience all the phases, with certain physicians even proceeding straight to TRT without establishing a specific diagnosis (see [Figure 15](#)). However, the synthesis also highlighted the lack of high-quality evidence across a range of populations.

Across included studies, the symptoms were discussed more often than the condition per se (hypogonadism), and these symptoms seem to affect men's daily lives in different ways. Similarly, the perceived effects of testosterone therapy were mainly discussed in the light of experienced symptoms. For instance, sexual desire or activity was the symptom more frequently discussed by participants,

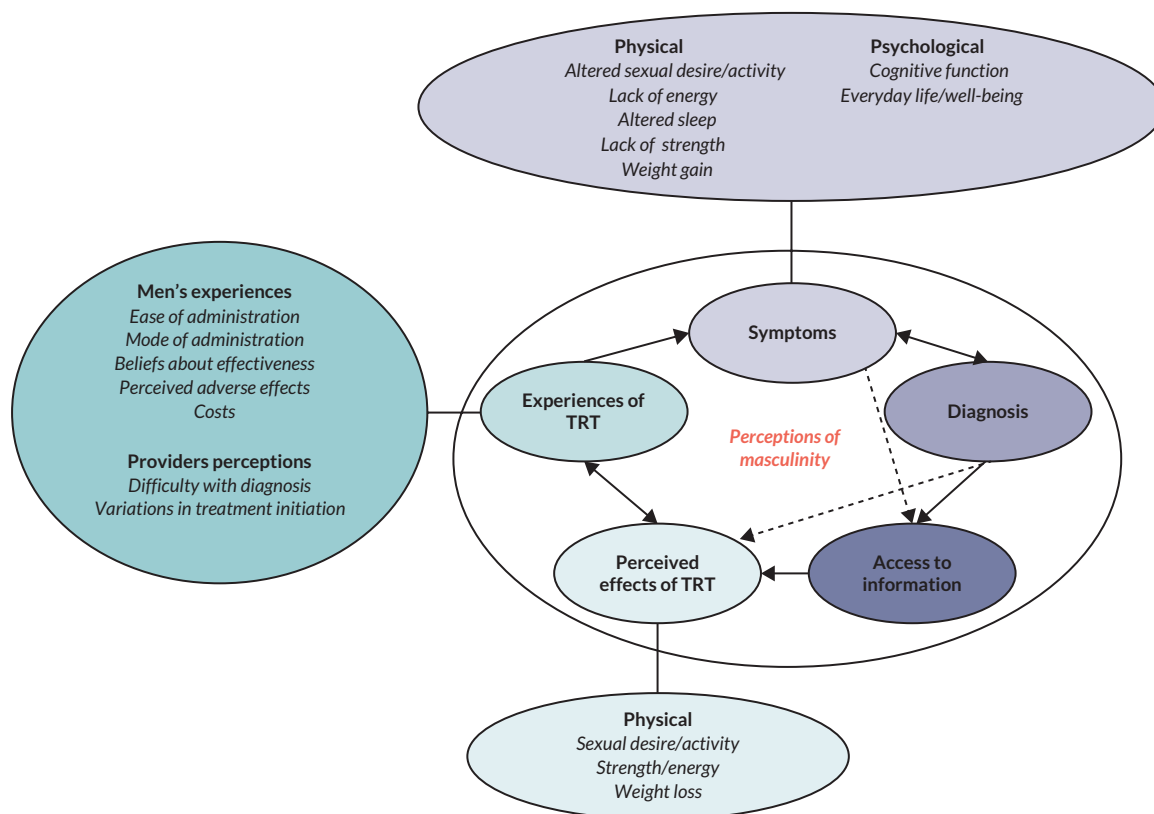


FIGURE 15 Conceptual diagram of the evidence on men and their healthcare providers' experiences of low testosterone and TRT.

not only as a symptom of low testosterone but also as a trigger for seeking professional advice, or an outcome expected to improve with TRT. Furthermore, sexual desire/activity was by far the most reported subtheme in two analytical themes (i.e. symptoms of low testosterone and impact on daily life, and perceived effects of TRT). It is unclear from the observed data whether the frequent reporting of sexual desire/activity reflects a reality of patients' experiences or is an artefact of the authors' study design or a responder bias.

Furthermore, some participants across the included studies did not attribute symptoms exclusively to low testosterone but also to ageing. Reported symptoms differed between treated and untreated hypogonadal participants. For instance, the control participants (men with normal levels of testosterone) in the Rosen 2009 study reported some of the recurrent themes discussed by men with low levels of testosterone (e.g. sleeping problems, cognitive function symptoms like memory loss or low concentration). Interestingly, the authors acknowledged that the subjective reports of specific symptoms (e.g. fatigue, decreased concentration, sexual desire/activity, sleep problems) might challenge the attribution of these symptoms specifically to hypogonadism. However, they explained that in their study, the control participants attributed these symptoms to other clinical conditions or health problems (e.g. weight gain because of overeating). Some participants with hypogonadism also associated some symptoms with other experienced symptoms. For example, some participants linked affective and emotional symptoms with sleeping problems and fatigue.

The lack of a standard definition and criteria to diagnose hypogonadism was acknowledged by the health professionals interviewed by Mascarenha *et al.*²⁸ In general, the perceived symptoms of low testosterone varied between participants, with the relationship between these symptoms remaining unclear. For instance, Rosen *et al.* suggest that the anxiety and depression reported by some of the hypogonadal participants were linked to their difficulties with sexual function.⁹ However, it is worth noting that this study showed that symptoms of anxiety and depression were also reported by those included in the control group who did not have low testosterone (men with normal testosterone).

To our knowledge, this is the first evidence synthesis focusing on the experience of men with hypogonadism and their acceptability of TRT, and on the views of health professionals who prescribe TRT. Our findings can be interpreted in the light of the results of other qualitative syntheses published in the literature that assess the impact of the context and treatment on specific symptoms related to hypogonadism (e.g. erectile dysfunction)^{85,86} or focus on participants with different clinical conditions (e.g. prostate cancer):^{85,87,88} for instance, the impact of symptoms on self-perception and masculinity. There is an indication that diminished perceptions of masculinity are a prominent concern among prostate cancer survivors, who might be more likely to be distressed by their erectile dysfunction and impact on their personal relations (partners and health professionals).⁸⁸ Perceptions of masculinity have also been reported to influence men's help-seeking behaviour for depression, with conformity to traditional masculine norms influencing access to and engagement with care.⁸⁹ However, in the context of hypogonadism and TRT, the overall effect of low levels of testosterone and experienced symptoms need to be further investigated. There is also a need to ensure research includes populations whom the evidence is required to serve. In other words, where it was reported, most men in the included studies were white. Ensuring that the concerns of men from other ethnic backgrounds (and indeed men with other protected characteristics such as sexual orientation, disability, etc.) are similar (or not) to the men included in this synthesis is also of critical importance if we are to ensure our findings are inclusive.

As reported previously, some examples of men's and physicians' behaviour described in these studies would undoubtedly lead to unnecessary prescribing of TRT:^{13,90} for instance, the described 'testosterone-seeking' attitude (wherein men were so determined to access TRT that they repeatedly sought new medical opinions until one eventually agreed to prescribe), along with the tendency of certain physicians to ascribe a broad generality of symptoms to 'low testosterone' and thus prescribe TRT with no prior meaningful diagnostics. This qualitative synthesis also highlights how physician knowledge, experience and preferences may impact upon the extent to which men might ascribe their symptoms to low

testosterone level (or make alternative associations) and, hence, affect their expectations of what TRT might realistically achieve for them. Furthermore, data from the current study suggest that men with hypogonadism experience an unmet need to receive a coherent, holistic narrative of their condition from their physicians that is not broken down into disconnected chunks labelled 'sexual function', 'mental health' or 'physical performance'.

Limitations

We need to acknowledge some limitations in the conduct of this evidence synthesis. Information on frequency of symptoms and characteristics of TRT (i.e. type, dose, route of administration, frequency of use) was poorly reported across included studies. Specific conditions (e.g. Klinefelter syndrome, congenital hypogonadism, prostate cancer, etc.) may lead to symptoms of hypogonadism itself alongside several unrelated symptoms. We therefore excluded any study restricted to a single aetiology of hypogonadism. While the approach taken in our analysis led to a small number of included studies, loosening the inclusion criteria may have paradoxically weakened our conclusions. Furthermore, as all studies were conducted in North America, the generalisability of their results to other health systems and social contexts, such as the UK, is questionable. More country-specific research in this clinical area would be useful. Most of the studies reported quotes derived directly from the participants to support identification of specific theme/subtheme; however, some studies provided only the interpretation of the authors. All the included studies assessed the perspectives of participants and only one study assessed the view of health providers in terms of barriers or facilitators to prescribing of TRT. The quality of the five included studies according to the CASP tool showed that the results across studies were valid and relevant to the scope of this synthesis. However, the small number of identified studies is an inevitable limitation of this work.

Conclusions

Overall, there is a paucity of qualitative evidence on the effects of low testosterone and consequences of TRT when compared with the number of existing clinical trials assessing the effectiveness of TRT. Our results indicate that the effects of low testosterone and subsequent treatment with TRT have multiple impacts and concerns for men. Many of the direct physical impacts of low testosterone (such as altered sexual desire and activity) appear to have multiple knock-on effects on other aspects of well-being, such as perceptions of masculinity and self-esteem; these further impact on the broad experience of everyday life. Therefore, it is critical to tackle these areas of concern effectively.

Based on this qualitative evidence synthesis, we make three recommendations. Firstly, some men with hypogonadism may benefit from a holistic, patient-centred approach to improving well-being and QoL, rather than the traditional focus on discrete symptoms (often sexual) practised by most clinicians; clinical psychologists, dietetics and physical therapists may help men for whom TRT is either not indicated or has not improved symptoms impacting on QoL. Secondly, the experience of men with hypogonadism is likely to be profoundly influenced by cultural identity and background, but our study reveals that this hypothesis remains unexplored; studying the impact of hypogonadism on men within different populations could improve the targeting of information and treatment monitoring for under-served demographic groups.⁹¹ Finally, clinicians need more support giving men with symptoms attributable (or not) to hypogonadism access to unbiased, patient-focused educational resources. Such future approaches would have a major, positive impact on the quality of health care for men with hypogonadism.

Chapter 4 Disease-specific patient-reported outcome measures for low testosterone: an investigation of item content to assess conceptual comparability

Introduction

To better understand the impact of low testosterone (and its treatments) on men with hypogonadism, combining the clinical outcome data on the safety and efficacy of these treatments with the patient-reported outcomes when evaluating in clinical trials provides a more holistic representation of the impacts. Patient-reported outcomes are often assessed and collected using questionnaires called PROMs. PROMs have been developed across many different disciplines and for a range of purposes (e.g. generic QoL measures compared to disease-specific). PROMs are defined as measures that are reported directly from the patient without modification or interpretation by a clinician or anyone else.⁹² Individual PROMs can be made up of numerous items and scales that (should) assess outcomes that matter to patients. However, the multiplicity of items and scales, for both generic and disease-specific PROMs, can result in the combining and aggregation of dissimilar items and scales that on face value appear to be measuring similar concepts. For example, two disease-specific measures report to evaluate QoL, but the individual items within those measures exhibit variation across core constructs.

When considering PROMs for patients with low testosterone, Langham *et al.* 2008 systematically reviewed health-related quality-of-life (HRQoL) instruments used in studies of adult men with testosterone deficiency to critically assess whether they accurately measure patients' concerns from the perspective of the tools' measurement properties.⁹³ The study identified 29 articles that included 14 HRQoL questionnaires (10 generic and four disease-specific) from 20 intervention studies, seven studies of the impact of low testosterone on the patient, and two studies describing the development of a HRQoL tool. Overall, this review found that PROMs used to measure HRQoL show changing measurement properties and often lack adequate clinical face validity.⁹³ However, this study did not explore whether these patient-reported HRQoL tools (specifically the disease-specific) are conceptually similar at an individual item level, raising a broader question about whether and how they can be meaningfully aggregated together in meta-analyses. In addition, the involvement of patients in the development of the disease-specific PROMs was not reported in this review, which is important when considering if the PROM adequately captures the concerns of outcomes that matter to patients.

In order to address these gaps, we conducted a systematic narrative review to:

- critically appraise the evidence on the item content of validated disease-specific PROMs for low testosterone or hypogonadism
- identify core domains of potential importance in this context.

Methods

Search methods for identification of studies

We applied three strategies to identify relevant disease-specific PROMs. First, we used an existing review on 'Health-Related Quality of Life Instruments in Studies of Adult Men with Testosterone Deficiency Syndrome' to identify relevant disease-specific PROMs.⁹³ This was supplemented by

searching for any relevant titles or abstracts identified in the systematic search conducted for the qualitative evidence synthesis (see [Chapter 3](#)). The rationale was that this search could identify papers that report both the development and validation of disease-specific PROMs. We also reference-chained by checking references of included studies to identify additional relevant PROMs. Finally, we identified and included any disease-specific PROMs that had been reported in the list of outcomes in the trials included in the individual patients' data (IPD) meta-analysis (see [Chapter 2](#)).

Inclusion criteria

1. Measures should aim to measure any patient-reported aspects of the impact on men's lives of living with low testosterone.
2. Study sample should enrol adult men (> 18 years old) with a diagnosis of hypogonadism confirmed by low levels of serum testosterone.
3. Study should include a validated questionnaire-based measure (to include self-report, administered by interview or by proxy).
4. Study should be published as a full-text original article.
5. Articles in English language will be included.

Exclusion criteria

1. Studies that reported study-specific measures with no details on development or validation.
2. Studies that included men with hypogonadism due to genetic or congenital causes and to concurrent clinical conditions (e.g. Klinefelter syndrome, congenital hypogonadism, prostate cancer, etc.).

Study selection and data extraction

One researcher (MA-M) independently screened all titles and abstracts identified in the systematic search with a randomly selected 10% checked by a second researcher (KG). When consensus could not be reached regarding eligibility, a discussion took place at a research team meeting and was resolved with a third reviewer (JC). We included any primary studies that explored any aspect of living with low testosterone, symptoms and/or treatment. Only PROMs available in English were considered for this review.

We developed PROM data extraction forms and data tables for each stage of the extraction process to standardise the information recorded and aid analysis. All data extracted and presented relate to data about the study that reported the development and validation of a measure to collect disease-specific PROMs for hypogonadal men. Data were extracted by one reviewer (MA-M), who recorded the name of the PROM(s), the reported PRO scales and individual verbatim items. We also extracted data on study characteristics such as country, population (including testosterone levels), sample size, response rate, age, ethnicity, employment status and education status.

Coding relating specifically to content items (i.e. individual questions) of the measure was conducted by two reviewers (MA-M and KG) independently, with any disagreements resolved through discussion. Data extraction was based on the following categories: name of the PROM; the concept of PROM (verbatim from the included study); items (verbatim from the included study); construct being targeted by item (verbatim from the included study); domain (as defined by review authors and informed by construct targeted).

While all PROM items were recorded to report PROM characteristics, only items measuring an experiential aspect of low testosterone were coded in the domain analysis. For example, questions relating to diagnostic criteria (such as onset of puberty or whether they have pituitary disease) were not domain coded. Equally we excluded experience follow-up questions that asked about how impactful the issue was (e.g. 'Do you experience morning erections? If yes, how often?'). In addition, in one of the PROMs (HIS-Q), a short version of the measure was also identified. To avoid the over-reporting of items from this short version that were already included in the original, only items that differed from the original version were extracted and included in the analysis ($n = 3$).

Data synthesis

Previous studies that have analysed PROMs into individual outcome domains informed analysis.⁹⁴⁻⁹⁶ We analysed the individual verbatim items from each PROM using a directed content analysis approach, which uses existing theory or research to identify key concepts as the initial coding categories.⁹⁷ The first step involved coding all of the individual items identified from each included PROM into domains defined by the review team. All PROM items were systematically categorised into conceptual 'health domains' (i.e. coding categories) according to the overall concept they aimed to capture; however, items were coded to more than one domain where appropriate. The coding categories were informed by the domains reported by how the PROM developers had reported the underlying construct and also further supplemented by existing classifications of health. The classification that we used to further define relevant health domains was the World Health Organization International Classification of Functioning, Disability and Health (WHO-ICF).⁹⁸ All individual items were extracted and mapped to a health domain, which was primarily informed by how the authors reported the underlying construct in the included study. Domain mapping was conducted by two reviewers (MA-M and KG) independently, with any conflicts resolved through discussion.

Both for descriptions of included studies and the PROMs reported in those studies, descriptive statistics were used to describe general information and measure detail. A narrative synthesis of the PROM and their inter-related domains is presented.

Results

The database search identified 1365 abstracts. After title and abstract screening, five PROMs (presented in six publications) met our inclusion criteria and were retrieved for further assessment. One PROM (HIS-Q) was also identified in a short form version. An additional three PROMs (presented in seven publications) were identified after screening references. No additional disease-specific PROMs were identified from the trials included in the IPD. Therefore, a total of 9 PROMs (presented in 13 studies) measuring experiences of men with low testosterone were included in this review (see [Figure 16](#)). The PROMs identified were: ADAMs Questionnaire,⁹⁹ the AMS scale,¹⁰⁰⁻¹⁰³ ANDROTEST©,¹⁰⁴ the age-related hormone deficiency dependent quality of life questionnaire (A-RHDQoL)©,¹⁰⁵ HED,⁸¹ hypogonadism impact of symptoms questionnaire (HIS-Q),^{82,106} HIS-Q-Short Form (HIS-Q-SF),⁸³ Massachusetts Male Aging Study (MMAS) questionnaire¹⁰⁷ and SAID.⁸¹ Two PROMs (HED and SAID) were described within the same study.⁸¹

Descriptive characteristics: included studies

We present the main characteristics of the included studies in [Table 16](#). The included studies were set within North America (USA $n = 5$ and Canada $n = 1$) or Europe (UK $n = 1$, Germany $n = 1$, Italy $n = 1$) and published between 1999 and 2016. Overall, the methods reported to develop the PROMs varied across studies and included surveys and interviews with samples ranging from 35 to 879 participants. The reported age of men included in the studies ranged from 18 to 85, with only two studies reporting mean values. Other demographic characteristics (ethnicity, sexual orientation, employment, education) of men in the included studies were often not reported (see [Table 16](#)).

Only two studies reported including men in the development of the PROM at the stage of identifying which items should be included in the PROM.^{81,82} The remaining seven PROMs were reported to be designed by a team of health professionals or were based on previous PROMs measuring general health-related outcomes. Most of the items were designed to measure symptoms of hypogonadism, with only one measuring changes before and after androgen replacement therapy (see [Table 16](#)).¹⁰¹

Descriptive characteristics: PROMs from included studies

Information relating to characteristics of the PROMs identified in the included studies is presented in [Table 17](#). The included studies report nine separate PROMs whose aim is to measure any aspect of

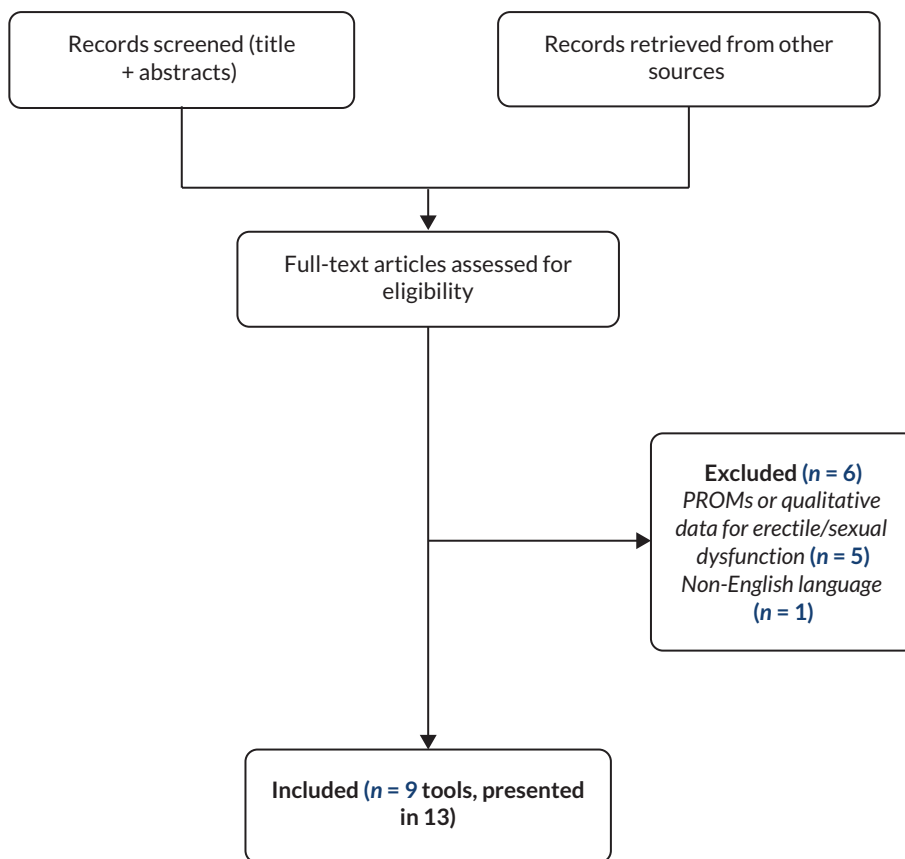


FIGURE 16 PRISMA flow diagram for identification of the PROM studies.

TABLE 16 Main characteristics of PROM studies

PROM Country Ref ID	PROM general characteristics	Sample size	Participant characteristics	Study methods
ADAM Questionnaire Canada Morley 2000 ⁹⁹	A tool created based on the author's clinical experience, 10 symptoms were identified and used to create this tool.	350	316 Canadian male physicians aged 40–82 years (mean 52.8) 34 male patients tested at the Saint Louis University Sexual Dysfunction Clinic, from which 21 had low bioavailable testosterone levels and received treatment with 200 mg testosterone cypionate intramuscularly every 2 weeks for 3 to 4 months. Treated 15 hypogonadal men (BT < 70 ng/dl) and 6 men with significant symptoms on the ADAM questionnaire and borderline gonadal function (BT ≤ 85 ng/dl) with testosterone cypionate 200 mg intramuscularly every 2 weeks. No information reported on time since diagnosis.	Participants completed the questionnaire and provided a serum sample for the measurement of TT, bioavailable testosterone and luteinising hormone. A group of 10 males completed the ADAM questionnaire on two occasions 2 to 4 weeks apart to determine the coefficient of variation for the questionnaire.

TABLE 16 Main characteristics of PROM studies (continued)

PROM Country Ref ID	PROM general characteristics	Sample size	Participant characteristics	Study methods
The AMS scale Germany Heinemann 1999 ¹⁰³ Heinemann 2001 ¹⁰² Heinemann 2002 ¹⁰⁰ Heinemann 2003 ¹⁰¹	Tool created to assess symptoms and their severity, and to measure changes before and after androgen replacement therapy.	116	Males (aged over 40 years) were recruited to complete a questionnaire of symptoms in seven practices of the ambulatory medical service in Berlin. As a first task, all patients completed the draft symptom inventory. The eligibility of male patients for androgen therapy was determined by the prescribing urologist, that is, following the recommendations of the International Society for the Study of the Aging Male (ISSAM) 11 nmol/l. ISSAM recommends that it is not yet known what level of serum testosterone defines deficiency in an older man, although it is generally accepted that 2 SD below normal values for young men is conclusively abnormal (11 nmol/l TT or 0.255 nmol/l free testosterone). For bioavailable testosterone the value of 3.8 nmol/l has been recommended. ¹⁰⁸	The development of the scale started with a comparison of over 200 variables in more than 100 medically well-characterised males. A factorial analysis was applied to establish the raw scale of complaints or symptoms that are not related to diseases, treatment, social and other variables, but related to ageing. The English translation included five experienced bilingual translators and reviewers. Troubles of compatibility between the cultural backgrounds of Germany, the UK and North America were identified and resolved by consensus. This resulted in one version for British and American English. This tool has been translated and culturally adapted into 12 languages.
ANDROTEST© Italy Corona 2006 ¹⁰⁴	Tool created by a team of andrologists, endocrinologists and psychologists, who were part of the clinical staff of an andrology unit. Each of the relevant areas identified was investigated through a few specific questions. The resulting interview was composed of 80 items. An exploratory analysis was performed to assess the association of scores of individual items with low testosterone. Those items which did not show any significant association with low testosterone were excluded.	879	215 male patients attending an outpatient clinic for sexual dysfunction from January 2002 to January 2003 and a further 664 patients were enrolled from February 2003 to February 2006. Patients with intellectual disability, or not fluent in Italian, were excluded. Results presented per sample. The mean age was 54 years for both samples. Most interviewed men were married or living with a partner. Men with diabetes who were also hypogonadal. Hypogonadism was defined as circulating TT below 10.4 nmol/l (300 ng/dl). No information reported on time since diagnosis.	Tool administered by interview. Patients were interviewed before the beginning of any treatment, and before any specific diagnostic procedure by two of the authors.

continued

TABLE 16 Main characteristics of PROM studies (continued)

PROM Country Ref ID	PROM general characteristics	Sample size	Participant characteristics	Study methods
The A-RHDQoL© UK Bradley 2001 ¹⁰⁹	Designed to measure the QoL of older men with age-related hormonal decline. This tool was influenced by previous tools (i.e. Schedule for the Evaluation of Individual Quality of Life, the Audit of Diabetes-Dependent Quality of Life and subsequent adaptations for people with macular degeneration, and with adult GHD).	128	Men being screened to include in a trial of GH and TRT. Participants had a mean age of 70.2 years, and the mean age of leaving full-time education was 17.3 years. Three participants reported stable disabilities. 32.8% of the participants reported no illnesses. The participants were older men who were being screened for inclusion in a trial of GH and TRT. The inclusion criteria were male sex, and age 65–80 years, i.e. no selection for hypogonadism. Unclear from paper, but appears that some of the included participants were excluded from the RCT because they were not hypogonadal. Mean TT 15.6 nmol/l in participants in the analysis. No information reported on time since diagnosis.	Participants had to answer the questionnaire. One of the 129 returned questionnaires had no items completed, and this was not included in analyses.
HED and SAID USA Hayes 2015 ⁸¹	Tool designed to assess the level of sex drive in men with hypogonadism. This study describes the creation of HED and SAID tool.	125	Men between age 18 and 85 years. For Studies A and B, eligibility included the diagnosis of hypogonadism. For Study C, eligibility included a clinically documented diagnosis of early-onset or congenital hypogonadism and two readings in the participants' medical charts of TT levels < 300 ng/dl. For Study D, eligibility included that of Study C, with the exception that participants were required to have a diagnosis of hypogonadism but did not have early onset. Studies 1 and 2, eligibility criteria included a written verification of a diagnosis of hypogonadism [either a prescription for low testosterone treatment or a laboratory sheet indicating a TT level < 300 ng/dl (10.4 nmol/l)]. For Study 3, eligibility criteria included a clinically documented diagnosis of early-onset or congenital hypogonadism and two readings in the participants' medical charts of TT levels < 300 ng/dl (10.4 nmol/l). For Study 4, eligibility criteria were similar to those of Study 3, with the exception that participants were required to have a diagnosis of hypogonadism but did not necessarily have early onset. No information reported on time since diagnosis.	Four separate qualitative studies were conducted. The same interviewer conducted all focus groups (Study A only), individual in-depth interviews (Studies A and C) and cognitive interviews (Studies B and D). (A) focus groups/interviews to identify important concepts related to the experience of hypogonadism and its treatment in men primarily with adult-onset hypogonadism. (B) Tested items generated for measurements of low sex drive and low energy. (C) Used interviews to confirm in men with early-onset hypogonadism that low sex drive and low energy were also essential symptoms. (D) Tested final versions of the two PROs and determining equivalency of paper-based and electronic versions of the two PROMs.

TABLE 16 Main characteristics of PROM studies (continued)

PROM Country Ref ID	PROM general characteristics	Sample size	Participant characteristics	Study methods
HIS-Q USA Gelhorn 2016 ^{82,106}	The development of this tool included a literature review, input from experienced clinicians, and qualitative interviews/focus groups with 65 male participants.	65	Men, age > 18 years; with hypogonadism [serum TT level < 300 ng/dl, (10.4 nmol/l)] and ability to read, speak and understand English. Participants in the first two phases of the qualitative part of the tool creation either were on androgen therapy, were not currently on androgen therapy or had been on androgen therapy for < 6 months. Participants in all phases were excluded if they had a significant health issue. Mean time since diagnosis (clinic report), years (SD) [range] 2.8 (2.2).	The first phase included focus groups and individual telephone interviews for concept elicitation (n = 30) and subsequent cognitive interviewing (n = 21) on the draft instrument. Cognitive interviews are qualitative interviews for assessing patients' understanding and the acceptability of a draft instrument. Based on feedback from regulatory agencies, additional in-person concept elicitation interviews (n = 6) were conducted in the second phase with younger patients, with lower testosterone levels, not on testosterone therapy or ED medications, and with an average BMI. The third phase included additional cognitive interviews (n = 8), including three participants with congenital hypogonadism who completed the concept elicitation and cognitive components of the interview.
HIS-Q-SF USA Gelhorn 2016 ⁸³	Using the original version, a shorter version was created and tested.	35	Men 18–65 years of age (mean 53.2); a history of signs and symptoms consistent with a diagnosis of hypogonadism (< 300 ng/dl) able to read, speak and understand English. Participants recruited through three clinical sites located in New Jersey, New York and Washington State from November 2013 through November 2014. Most of the participants were white (71.4%). The mean time since hypogonadism diagnosis was 2.9 ± 3.9 years; 91.4% of participants had previously received TRT, and 88.6% were currently receiving TRT. The participants' mean BMI was 31.3 ± 5.4. Mean time since diagnosis (clinic report), years (SD) 2.8 (2.2).	Twenty men participated in concept elicitation through three focus groups (n = 18) and two interviews (n = 2), and cognitive interview (n = 15). A two-part qualitative study involving semistructured focus groups and one-on-one interviews and cognitive interviews. The first part included concept elicitation focus groups and discussions to solicit spontaneous input on patients' hypogonadism experiences, including sorting hypogonadism symptoms in order of importance. The second part involved cognitive interviews on the newly created draft version of HIS-Q-SF that focused on participants' understanding of the items, decision processes about the responses, interpretation of response options, and understanding and testing recall period appropriateness. The study included in-person and telephone discussions. Participants took part in only one stage of the research process (either the first or second part).

continued

TABLE 16 Main characteristics of PROM studies (continued)

PROM Country Ref ID	PROM general characteristics	Sample size	Participant characteristics	Study methods
MMAS questionnaire USA Smith 2000 ¹⁰⁷	Data for the screener construction phase were drawn from the MMAS. Men were interviewed in their homes by a trained interviewer-phlebotomist. Variables were chosen to represent the subject's present condition (e.g. symptoms in the last 2 weeks).	304	Tool validation was done in men aged 40–79 years presenting at a Massachusetts primary health-care clinic for routine check-ups, influenza vaccinations and minor medical problem. The field validation sample comprised 304 men who had complete, correctly filled out screening instruments, ancillary data on medical conditions, treatments and sexual function, and an available serum testosterone measurement. Testosterone deficiency was defined as serum TT level below 12.1 nmol/l (349 ng/dl). The study tested the MMAS questionnaire on men with and without low testosterone and reported the sensitivity and specificity to predict serum testosterone. No information reported on time since diagnosis.	The outcome of interest for the screener was testosterone deficiency. A brief mail survey of the Endocrine Society was done. Based on 53 responses, testosterone deficiency was defined as serum TT below 12.1 nmol/l. Also, potential items predicting testosterone deficiency were selected from the MMAS. The pool was restricted to questions that were clear in a self-administered instrument, and simple enough to answer in multiple-choice mode. Variables were chosen to represent the subject's present condition (e.g. symptoms in the last 2 weeks). Thirty-four variables met these criteria and were considered including in the screener. All candidate variables were dichotomised by selecting the grouping of levels that gave the largest OR and the greatest statistical significance for predicting testosterone deficiency. A final set of eight were independent, statistically significant predictors of testosterone deficiency. The resulting eight-item screening instrument was evaluated under the Receiving Operator Characteristic analysis. Then, a field test to assess the screener's validity in an independent sample was carried out. Patients were asked to complete and score a pencil-and-paper version of the screener without the help of clinic staff.

hypogonadism and related symptoms. The number of relevant experience-based items varied across PROMs and ranged from 3 to 53 items (median = 7) with a cumulative total of 98 individual items across the nine PROMs.

The response options and ease of scoring and administration varied across PROMs. Some of the PROMs ($n = 2$) included dichotomous response options, including a yes or no option.^{99,107} However, some others ($n = 5$) used questions evaluated with a Likert scale where the score increases or decreases in line with the perceived symptom severity.^{82,83,103–106} Four of the PROMs were patient self-completion PROMs, and the rest required a researcher or trained interviewer to recall and evaluate the patient's responses. Only one study reported the time the PROM took to complete (< 10 minutes).¹⁰⁴

TABLE 17 Characteristics of included PROMs

Tool Country Ref ID	Dimension (number of items)	Response options (range)	Ease of scoring and administration (range of scores)	Mode of administration	Sample items
ADAM Questionnaire Canada Morley 2000 ⁹⁹	10 items	Dichotomous.	Yes/no options	Self- completion.	Do you have a decrease in libido (sex drive)?
The AMS scale Germany Heinemann 2003 ¹⁰¹	17 items	5-point Likert scale.	1 = none, 2 = mild, 3 = moderate, 4 = severe, 5 = extremely severe. The score increases point by point with increasing perceived symptom severity.	Self- completion.	Decrease in ability/ frequency to perform sexually.
ANDROTEST© Italy Corona 2006 ¹⁰⁴	12 items	Open-ended questions, interviewer needs to grade based on a 3-point Likert scale.	Interview questions varied. The interviewer must score depending on the answer. The score increases point by point with increasing perceived symptom severity.	Verbally administered by a researcher.	Describe what happens during sexual intercourse: how often do you have lack of an erection?
The A-RHDQoL© UK Bradley 2001 ¹⁰⁹	50 items	Open-ended questions, with a 4- or 7-point Likert scale. Some questions have the option to answer 'Not applicable'.	For questions assessing the PROM a 7-point Likert scoring scale goes from very much better to very much worse. (The score decreases point by point with decreas- ing perceived symptom severity.) For each of the PROM questions, there is a second question assessing how important that PROM is in the man's life and there is a 4-point Likert scale that goes from 'very important' to 'not at all important'.	Self- completion.	If my hormone levels had not declined with age, my sex life would be: This aspect of my life is:
HED USA Hayes 2015 ⁸¹	3 items	Interview with open- ended questions.	Quotes are recalled.	Verbally administered by a researcher.	Right now, how tired or exhausted do you feel?
HIS-Q USA Gelhorn 2016 ⁸²	53 items	Interview with open- ended questions with 5-point Likert scale.	Scores ranged from 'not at all' to 'extremely' and from 'never' to 'always' for severity and frequency items.	One-on-one cognitive interviews.	Did you feel sexual desire? (Over the past 14 days ...)
HIS-Q-SF USA Gelhorn 2016 ⁸³	17 items	Interview with open- ended questions with 5-point Likert scale.	Scores ranged from 'not at all' to 'extremely' and from 'never' to 'always' for severity and frequency items.	One-on-one cognitive interviews.	How many times did you engage in sexual activities? (Over the past 14 days ...)

continued

TABLE 17 Characteristics of included PROMs (continued)

Tool Country Ref ID	Dimension (number of items)	Response options (range)	Ease of scoring and administration (range of scores)	Mode of administration	Sample items
MMAS questionnaire USA Smith 2000 ¹⁰⁷	8 items	Dichotomous.	Yes/No options. Some questions have specific dichotomous options.	Self- completion.	How much do you usually sleep? (< 5 hours OR 5 hours or more)
SAID USA Hayes 2015 ⁸¹	5 items	Interview with open-ended questions:	Quotes are recalled.	Verbally administered by a researcher.	During the past 7 days to what extent did you think about sexual activity?

Item domain classification

Our review identified 10 relevant health domains across the 98 items from the nine PROMs. Ten domains were defined according to the WHO-ICF classifications and were identified as: Cognition, Energy, General well-being, Mood, Pain, Physical – general, Role, Sexual, Sleep, Social. [Table 18](#) presents

TABLE 18 Domain definitions

Domain	Domain defined by WHO-ICF	Definition according to the WHO-ICF	Example of item coded into the domain
Cognition	b117 Intellectual functions	General mental functions, required to understand and constructively integrate the various mental functions, including all cognitive functions and their development over the life span.	'How well were you able to focus your attention on tasks?' Gelhorn 2016
Energy	b130 Energy and drive functions	Mental functions that produce vigour and stamina.	'Right now, how energetic do you feel?' Hayes 2015
General well-being	d570 Looking after one's health	Ensuring physical comfort, health and physical and mental well-being, such as by maintaining a balanced diet, and an appropriate level of physical activity, keeping warm or cool, avoiding harms to health, following safe sex practices, including such as using condoms, getting immunisations and regular physical examinations.	'Have you noticed a decreased enjoyment of life?' Morley 2000
Mood	b1263 Psychic stability	Mental functions that produce a personal disposition that is even-tempered, calm and composed, as contrasted to being irritable, worried, erratic and moody.	'Irritability (feeling aggressive, easily upset about little things, moody)' Heinemann 2003
Pain	b280 Sensation of pain	Sensation of unpleasant feeling indicating potential or actual damage to some body structure.	'If my hormone levels had not declined with age, my levels of body pain would be:' Bradley 2001, McMillan 2003
Role	d7203 Interacting according to social rules	Acting independently in social interactions and complying with social conventions governing one's role, position or other social status in interactions with others.	'Do you like directing other people's work?' Smith 2000
Sexual	b640 Sexual functions	Mental and physical functions related to the sexual act, including the arousal, preparatory, orgasmic and resolution stages.	'Have you had more or less desire to make love in the last 3 months?' Corona 2006
Sleep	b134 Sleep functions	General mental functions of periodic, reversible and selective physical and mental disengagement from one's immediate environment accompanied by characteristic physiological changes.	'How much difficulty did you have getting enough sleep at night?' Gelhorn 2016
Social	d710 Basic interpersonal interactions	Interacting with people in a contextually and socially appropriate manner, such as by showing consideration and esteem when appropriate, or responding to the feelings of others.	'If my hormone levels had not declined with age, my friendships and social life would be:' Bradley 2001, McMillan 2003
Physical – general	NA	^a	'Have you recently been bothered by headaches?' Smith 2000

^a In this item, any other physical symptom reported was recorded.

Note

Definitions retrieved from WHO-ICF webpage.⁹⁸

a summary of the domain labels, definitions, and example items from the PROMs. These 10 domains are conceptually distinct; however, many of them are interconnected, with changes in one influencing changes in others (e.g. sleep and energy).

The main domains identified across PROMs are presented in [Table 19](#). The most frequently identified domain across PROMs was the sexual domain, with 29 (29.6% of total items) items measuring this concept across the PROMs. However, two of the PROMs, HED and MMAS, did not include any items that covered the sexual domain.^{81,107} The next most frequently identified domains were mood and role, with 14 items each (14.3% of total items), followed by energy and sleep with ten each (10.4% of total), but again items measuring these domains were not consistently included in all PROMs. The domains that were least frequently identified were pain, with a total of only two items across two PROMs (2.1%), and the social domain, with three items which were all included the A-RHDQoL PROM.^{105,109}

Six of the nine PROMs can be considered as being multidimensional (i.e. capturing more than one domain) based on their coded domains, with three PROMs being unidimensional (i.e. only capturing one domain). The A-RHDQoL PROM showed itself to be the most comprehensive across the PROMs included since this was the only one to include items (26 items), which could be coded to all 10 domains and the only tool to include the social domain. Of the three PROMs which were unidimensional in their mapping of items to domains, all included items that were coded to the sexual domain.

Table 19 Domains identified across relevant individual items from the assessed PROMs

Domain (n = 10)	No. of items in the domain	ADAM	AMS	ANDROTEST	A-RHDQoL	HED	HIS-Q	HIS-Q-SF	MMAS	SAID
Cognition	4	0	0	0	2	0	2	0	0	0
Energy	10	2	2	0	2	2	2	0	0	0
General well-being	6	1	1	0	3	0	1	0	0	0
Mood	14	1	6	0	3	0	4	0	0	0
Pain	2	0	1	0	1	0	0	0	0	0
Physical – general	6	1	2	0	2	0	0	0	1	0
Role	14	2	1	0	7	0	3	0	1	0
Sexual	29	2	3	7	2	0	7	3	0	5
Sleep	10	1	2	0	1	1	4	0	1	0
Social	3	0	0	0	3	0	0	0	0	0
Total items per tool domain coded	98	10	18 ^a	7	26	3 ^a	23	3	3	5
Total domains per tool	NA	7	8	1	10	2	7	1	3	1

^a These tools included 1 item that was double coded to two separate domains.

Note

HIS-Q-SF was a short version of HIS-Q. For this reason, we included only items that were different between these tools in the analysis. In HIS-Q short version, three items test the sexual domain, which was covered in the original version, but the questions in the short form are different.

Discussion

This review of disease-specific PROMs for men with hypogonadism is the first to systematically characterise the item content across these measures into individual outcome domains. Our review highlights the heterogeneity that exists across these PROMs all reporting to capture QoL relating to the same disease. Previous publications of the measurement properties of health-related (generic and disease-specific) QoL instruments for low testosterone examined the clinical face validity of the PROMs. However, this previous review did not compare the content of items across PROMs to explore how similar, or not, these tools are. The present review extends these findings to highlight the heterogeneity in domains measured across PROMs, questioning the content validity of these disease-specific PROMs.

As shown earlier in the results, the PROMs identified varied in terms of the number of individual items (ranging from 3 to 53 items) and the domains covered by such items. The only PROM that covered all 10 domains defined in this review was the A-RHDQoL and this was the most extended PROM, with 26 questions testing the 10 identified domains.¹⁰⁹ Interestingly, this PROM concentrated most of its items in the role domain, where conversely many other PROMs items concentrated in the sexual domain. It may not be surprising that the PROM with the highest number of items also covers the largest number of domains. However, the HIS-Q includes 23 items, and is the second-longest, but only covers seven out of 10 domains. In contrast, the AMS contains fewer items ($n = 18$) but covers more ($n = 8$) domains. It is important to note that the variation in these PROMs raises a broader question about the adequacy of combining such conceptually divergent measures in summative assessments. For example, all PROMs report to cumulatively assess men's QoL, but if a meta-analysis combined data from the ANDROTEST and SAID PROMs it would only provide a measure of impacts on sexual function, but other concepts may also be important for overall assessments. Ensuring this is accounted for in interpretation of results of studies of this kind is important.

The sexual domain contained the highest number of items reported across all PROMs, with 30% of the included items. This is not unexpected given testosterone is a sex hormone and therefore this is an obvious outcome to be impacted and reported by men with low testosterone and hence its need to be included in the PROMs. However, other effects of low testosterone such as impacts to cognition are less well represented, making up only four items. Of note is that one PROM contributing half of these items is one of the four PROMs that had patient involvement in item conception (HIS-Q). Given symptoms experienced by patients with hypogonadism are often multifactorial and impacts on one can directly influence impacts on other aspects of QoL, considering whether and how existing PROMs adequately capture the multidimensional nature should be considered further. This raises questions about the adequacy of development and coverage of items that matter to men with low testosterone.

The findings identified a lack of input from men with low testosterone in PROM development, specifically during item conception and identification for inclusion, the critical phase to ensure patient-relevant outcomes are represented in the PROMs. Only HIS-Q (Gelhorn 2016),⁸² HIS-Q-SF (Gelhorn 2016b),⁸³ HED (Hayes 2015)⁸¹ and SAID (Hayes 2015)⁸¹ included patients while conceptualising and defining the relevant items to be considered in the PROMs. This may not be unexpected given these PROMs are amongst the most recently developed and therefore the importance of involving patients in PROM development (and research more generally) is much more widely accepted as a mechanism to ensure research is directly relevant to patients and produces more meaningful outputs.

Linked to this lack of involvement of men during PROM development is the question of whether all representative populations have been involved in the development of the PROMs. Our findings highlight the predominant lack of detail on the participants included in the studies developing these PROMs (e.g. ethnicity, sexual orientation). Similar to the findings from the qualitative evidence synthesis (See [Chapter 3](#)), it is important to consider how perceptions of masculinity might differ amongst different groups of men and ensuring these men are represented during development of a PROM to capture experience of low testosterone is critical. Future research should engage with men from a range of

populations during PROM development to ensure included outcomes represent those of the community they deem to serve.

Strengths and limitations

This review of PROMS for low testosterone forms part of a mixed-methods complex evidence synthesis project that included a detailed systematic search to identify qualitative evidence and disease-specific PROMs and included rigorous methods to identify and code relevant domains across included patient-reported measures. The coding of items was conducted by two authors independently. Directed content analysis was used to analyse the included items, with some level of interpretation required for coding the items. Therefore, while we applied a systematic and rigorous approach, like many qualitative interpretive approaches it is subjective, and it is possible that if conducted by other researchers (with different perspectives and lenses) a different overall result may emerge. Also, only PROMs in English were considered. We are aware that there are PROMs available in a language other than English (e.g. German – Hypogonadism Related Symptoms Scale from Wiltnik *et al.* 2009).¹¹⁰ However, as highlighted by Heinemann 2003, given there might be problems associated with compatibility between the cultural backgrounds and language translation, which might even vary within the same language (e.g. PROM translation consensus resulted in one version for British and American English), we chose to exclude these PROMs.¹⁰¹

Conclusions

This study has shown the considerable variability that exists in disease-specific PROMs for men with low testosterone regarding development and domain coverage. It has also highlighted the lack of input from men in the development of these PROMs, bringing into questions their relevance and adequacy in capturing outcomes that matter to men with low testosterone. The dominant focus of these PROMs to date has centred on sexual function, but possibly to the detriment of other aspects that also matter to patients.

Chapter 5 Cost-effectiveness of testosterone replacement therapy

The economic evidence on TRT was assessed through a systematic review of economic evaluations as well as a new model-based economic evaluation comparing TRT with standard of care (SoC; e.g. no treatment).

Systematic review of existing cost-effectiveness

A systematic literature review was conducted to identify economic evaluation studies assessing the use of TRT compared with alternative treatments in men with testosterone deficiency.

Inclusion and exclusion criteria

We focused on full economic evaluations reporting the cost and consequences of at least two alternative care pathways (i.e. TRT compared to 'standard care' – no treatment). No restrictions were imposed on the way costs and effects were calculated; therefore, cost–consequences, cost–effectiveness, cost–utility and cost–benefit analysis were deemed suitable for inclusion. Studies with hypogonadism caused by congenital disorders (e.g. Klinefelter syndrome) and participants with secondary hypogonadism were excluded unless study results were reported separately for the study population of interest (e.g. results for men with Klinefelter syndrome and results for men with hypogonadism).

Search strategy

Sensitive electronic literature searches using an appropriate combination of controlled vocabulary and text terms were developed. Relevant electronic databases [i.e. MEDLINE, Embase, NHS Economic Evaluations Database (NEED), the HTA Database, Cost-effectiveness Analysis Registry, and Research Papers in Economics] were searched from 1992 until 4 February 2021. Full details of the search strategies are reported in [Appendix 1](#). In addition, recent conference proceedings of key professional organisations in the fields of endocrinology (e.g. American Endocrine Society), cardiology (e.g. American College of Cardiology) and men's health (e.g. European Menopause and Andropause Society, International Society of Men's Health) for the last 3 years (2018–2020) as well as the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) Scientific Presentations Database were also scrutinised.

Study selection and data extraction

After electronic de-duplication, one reviewer screened titles and abstracts. All potentially relevant studies were retrieved for full-text assessment. One reviewer selected studies for inclusion. Any doubt about study selection was discussed with the members of the project team and the advisory group. Following current methodological standards, data from the included studies were extracted following the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) checklist.¹¹¹ The methodological quality of included studies was assessed using the checklist as critical appraisal questions and results were reported in a narrative manner with no attempt to synthesise them quantitatively.

Results of the systematic review of cost-effectiveness

After de-duplication, 454 abstracts were screened for suitability. Twenty-one studies were selected for full-text assessment and one study met our pre-specified inclusion criteria.

Arver *et al.* 2014 assessed the cost-effectiveness of testosterone undecanoate (TU) injection compared with no treatment in two patient populations: men with Klinefelter syndrome and men with LOH.¹¹²

Results for the two patient populations were reported separately and those related to LOH are considered in our review.

Arver *et al.* 2014 conducted a cost–utility analysis for Sweden using a patient-level simulation model that included the following health states: no complications, type-2 diabetes mellitus (T2DM), CV events because of T2DM (i.e. acute myocardial infarction and angina pectoris), CBV events because of T2DM (i.e. stroke and transient ischemic attack), major depression, fractures and death.¹¹² The authors used age-matched utility values for healthy Swedish males (range 0.91 for 20–29-year-olds to 0.74 for 80–88-year-olds) together with utility decrements due to events and complications (e.g. –0.121 for diabetes, –0.151 for depression). Cost of drugs, treatment administration and monitoring according to the Swedish healthcare provider were included. Indirect costs due to TRT treatment administration and complications were added. In addition, general population mortality risk for Sweden, the mortality risk for specific events (i.e. CV and CBV events) and subsequent conditions (i.e. increased age- and sex-dependent diabetes-specific mortality, depression and fractures) were considered in the model. The authors assumed normal testosterone levels achieved for all patients (100% treatment response rate). Moreover, based on two studies analysing retrospective medical records and one prospective cohort study, men under TRT benefited from a reduced risk of severe depression and a reduced risk of developing T2DM. Also, all men with T2DM in the model faced an increased risk of CV and CBV events. The model was run for a lifetime time horizon, costs were expressed in 2009 Euros and costs and effects were discounted at a 3% annual discount rate. The authors' results showed that TRT generated 1.13 additional quality-adjusted life-years (QALYs) per patient with an incremental cost of €22,229 compared with no treatment. The incremental cost per QALY gained was €19,720. Several one-way sensitivity analyses were conducted, with the model results being robust to all these. The authors concluded that the lifelong treatment with TU depot injection was a cost-effective treatment option for men diagnosed with hypogonadism in Sweden.

Arver *et al.* 2014 was the only economic evaluation that met our inclusion criteria.¹¹² While the study was informative a new model was developed to better reflect the findings of the clinical IPD analysis from the perspective of the UK NHS. Further discussion of the Arver *et al.* 2014 findings and a comparison with our model results are provided in the discussion to this chapter.

Model-based economic evaluation

Introduction

The objective of this analysis was to assess the cost-effectiveness of TRT compared with SoC. This cost-utility analysis was conducted following best practice in decision modelling. The model was developed in TreeAge Pro (Healthcare Version) Version 2021.¹¹³ A cohort Markov model incorporating relevant care pathways for individuals with low testosterone was informed by existing guidelines, the IPD meta-analysis and discussions with experts in the project advisory group. Utility data for individuals under TRT and SoC were also based on the analysis of the TestES IPD. In addition, the model considered care pathways for CV and CBV events. The model structure and strategies were agreed upon by the members of the Project Advisory Group (5 February 2021). For all the analyses we adopted an NHS and personal and social services perspective.¹¹⁴

Methods

Model structure

The schematic economic model structure is shown in [Figure 17](#). Adult men with hypogonadism enter one of the two model strategies, SoC or TRT. These strategies have similar model structures. All individuals start at the *No complications* health state and when they experience CV or CBV events, move

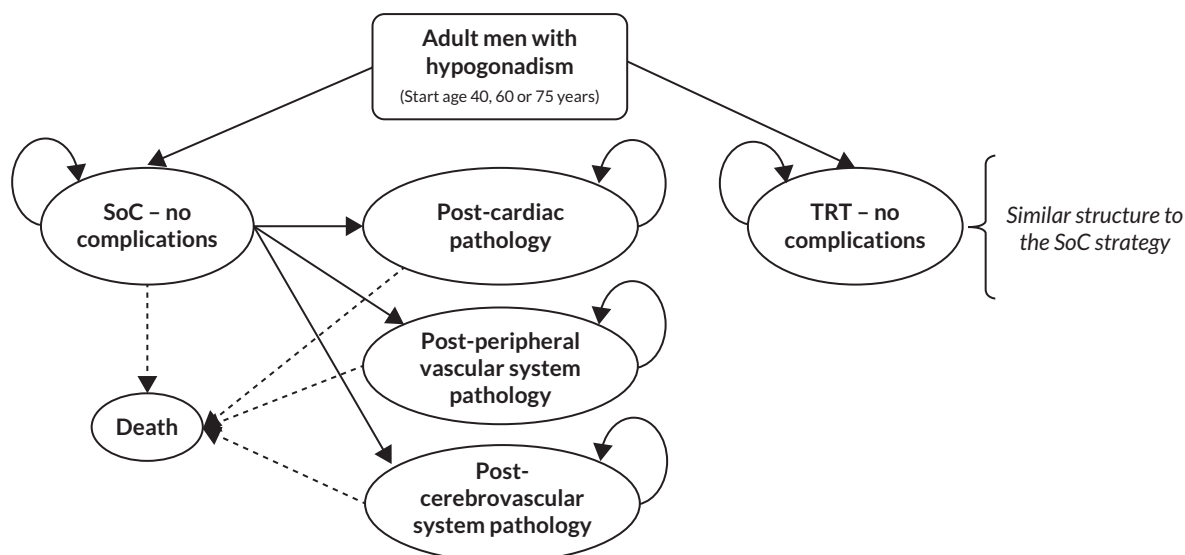


FIGURE 17 Simplified schematic of the economic model structure.

to the corresponding post-complication Markov state. The complications assessed in the model are the primary clinical outcomes of the IPD meta-analysis, which were organised into three main categories: cardiac pathology, pathology of the peripheral vascular system and pathology of the CBV system. The ‘*cardiac pathology*’ category included arrhythmia, CHD, heart failure, myocardial infarction, valvular heart disease, stable, new and unstable angina, aortic aneurysm and cardiac arrest; the ‘*pathology of the peripheral vascular system*’ category comprised peripheral vascular disease, aortic aneurysm, aortic dissection and atherosclerosis; and the ‘*cerebrovascular system pathology*’ category included stroke and transient ischemic attack. Mortality was considered with *Death* as an absorbing Markov health state. Markov cycle length was defined as 1 month.

Population

The economic analysis was conducted for a cohort of symptomatic men with testosterone deficiency. To illustrate the age groups of clinical interest (under 60 years old, 60–75 years old and over 75 years old), three starting age categories were defined: 40-, 60- and 75-year-old people.

Time horizon and discounting

A 10-year time horizon was chosen for the economic modelling. Given the 3-year follow-up of the RCTs included in the synthesis of clinical effectiveness evidence and the discussions with the advisory group for this project, we believed that extrapolation of clinical effects from the IPD analysis beyond 10 years would be highly uncertain. Nevertheless, given potential for long-term differences in complications, survival and costs, a sensitivity analysis was conducted to assess the impact on cost-effectiveness of applying a lifetime horizon as recommended by UK economic evaluation methods guidelines.¹¹⁴ A half-cycle correction was applied and future costs and QALYs were discounted at a rate of 3.5% per annum.¹¹⁴

Clinical input parameters

Primary outcomes from the analysis of TestES data were used as key model parameters (see [Table 20](#)). The RR for all-cause mortality and the risk of CV or CBV complications from the one-stage meta-analysis were incorporated in the model. Log-normal distributions were constructed based on the 95% CI (RR mortality: 0.47, 95% CI 0.18 to 1.25; RR complications: 1.06, 95% CI 0.82 to 1.38).

The number of events in the TestES data set was used to proportionally allocate cost and reduced utilities in the model. The numbers of events for each type of complication were pooled across study

TABLE 20 IPD meta-analysis outcomes incorporated into the economic model

Variable	Point estimate	RR 95% CI	Distributional form	Source
RR for any cause mortality	0.47	(0.18 to 1.25)	Log-normal	TestES
RR of CV and/or CBV complications	1.06	(0.82 to 1.38)	Log-normal	TestES

arms (see [Table 21](#)) to avoid artificially biasing the cost and/or utilities towards one arm due to non-statistically significant differences in the proportional distribution of event types between treatment arms in the IPD.

The underlying risk of experiencing a CV or CBV event was derived from the British Heart Foundation Heart and Circulatory Disease Statistics 2020.¹¹⁵ This compendium of statistics reports incidence rates per 100,000 adults per year for selected CV conditions (atrial fibrillation, heart failure, stroke, transient ischemic attack and peripheral vascular disease) by gender and age for the UK. The latest incidence rates for males available for 2017 were used and transformed into monthly probabilities (see [Table 22](#)) assuming a constant monthly rate.

Mortality

Age-specific mortality rates for males, sourced from the UK life tables, were used to model death from all causes in those with no complications.¹¹⁶ For those modelled to experience complications, the background mortality rate was adjusted using a standardised mortality ratio (SMR) to consider the higher risk of dying following the occurrence of complications. In search for these data, relevant UK clinical

TABLE 21 Proportion of CV and CBV complications

Type of event	Number of events (pooled)	(%)
Pooled total number of CV and/or CBV events	365	100%
Cardiac pathology		
Arrhythmia	99	27%
CHD	66	18%
Heart failure	50	14%
Myocardial infarction	26	7%
Valvular heart disease	30	8%
Stable angina	14	4%
New angina	10	3%
Unstable angina	6	2%
Cardiac arrest	2	1%
Pathology of peripheral vascular system		
Aortic dissection	2	1%
Aortic aneurysm ³	13	4%
Atherosclerosis (excluding CBV)	2	1%
Peripheral vascular disease	22	6%
Pathology of the CBV system		
CBV event (stroke or transient ischaemic attack)	23	6%

TABLE 22 Underlying monthly risk of CV events

Variable	Point estimate	Distributional form ^a	Source
45–54 years	0.0005	Beta: alpha = 553; beta = 99447	Circulatory Disease Statistics 2020. BHF
55–64 years	0.0011	Beta: alpha = 1318; beta = 98682	Circulatory Disease Statistics 2020. BHF
65–74 years	0.0020	Beta: alpha = 2335; beta = 97665	Circulatory Disease Statistics 2020. BHF
75+ years	0.0038	Beta: alpha = 4456; beta = 95544	Circulatory Disease Statistics 2020. BHF

^a Parameters correspond to yearly probability distributions; outcomes of these were transformed into monthly probabilities for the model.

guidelines and health technology assessments were reviewed as the development of these guidelines included systematic literature reviews that are specific to the clinical area of interest.

Standardised mortality ratio for the post-*Cardiac Pathology* and post-*Cerebrovascular System Pathology* health states were obtained from National Institute for Health and Care Excellence (NICE) guideline NG185 on Acute Coronary Syndromes.¹¹⁷ The SMR of 2.00 for the post-*Cardiac Pathology* health state was originally obtained from Smolina *et al.* 2012, a UK study that linked hospital and mortality data identifying 387,452 individuals in England who were admitted into hospital with an acute myocardial infarction diagnosis between 2004 and 2010 and were followed up for 7 years.¹¹⁸ SMRs of 4.73 and 2.32 were used for the post-*Cerebrovascular System Pathology* health state for the first and subsequent years, respectively. These were based on the original study by Bronnum-Hansen *et al.* 2001, which assessed 4162 individuals from the Copenhagen county region who suffered a brain stroke between 1982 and 1991.¹¹⁹ Finally, the SMR of 3.14 for the *Peripheral Vascular System Pathology* health state was retrieved from the NICE Clinical Guideline 147 and is based on Criqui *et al.* 1992, a relatively small study conducted in the USA that identified 67 individuals with large-vessel peripheral arterial disease (LV-PAD) who were followed, prospectively, for 10 years. The authors reported SMRs according to the presence or absence of LV-PAD.

Health state utilities

The TestES systematic literature review of RCTs identified the studies that collected QoL data. Special attention was given to those instruments from which direct measures of utility could be obtained such as EuroQol-5 dimensions (EQ-5D), SF-36 or SF-12. In addition, instruments used within the included studies related to sexual function, psychological function or QoL were checked against the Oxford database of mapping studies.¹²⁰ The Oxford database provides a readily available collection of all studies mapping patient-reported outcomes to the EQ-5D instrument. Full copies of the publications for the identified mapping algorithms were obtained with the aim of applying the mapping algorithms to estimate utilities for TRT and SoC groups using TestES IPD.

[Appendix 7, Table 46](#) shows the data available from the received data sets that allowed the direct estimation of utility scores. Five study data sets provided SF-36 individual item data; however, one of these only presented baseline data (Snyder *et al.* 2016),¹¹ leaving only data from four data sets for analysis. Data for baseline, 26 weeks (Basaria *et al.* 2015a, Emmelot-Vonk *et al.* 2008, Hildreth *et al.* 2013 and Magnussen *et al.* 2016),^{39,40,45,50} 52 weeks (Hildreth *et al.* 2013),⁴⁰ 78 weeks and 156 weeks (Basaria *et al.* 2015a)³⁹ were available. The 26-week data were regarded as the most robust as they were available for four studies, with the analysis conducted for $N = 409$ participants. The number of participants reduced substantially for later time points. Also, data on QoL at 1 year and beyond potentially reflect the reduction in QoL due to complications. The model accounts for utility reductions due to complications independently; therefore, there was a risk of double counting these utility reductions due to complications when using long-term trial data. Short form-6 dimensions (SF-6D) algorithms were used to estimate utility scores for each participant at baseline and 26 weeks.¹²¹ A mixed-effect regression model (random effects on study and fixed effects on participants) was run to estimate a difference of 0.0036 (95% CI -0.012 to 0.019) in utility score between TRT and SoC. The

regression predicted utility score was used for the SoC group (0.792) and the mean estimated difference was added to estimate the TRT score. Utility multipliers for TRT and SoC were calculated by dividing these utility scores by the population norm for the sample (i.e. 0.795 for 65- to 69-year-olds).¹²² Finally, these utility multipliers were applied to the general population EQ-5D score formula proposed by Ara and Brazier (2010) to obtain the age- and male-specific utility score for the *No complications* health state for TRT and SoC, respectively.¹²³

In addition to the SF-36 data, the IPD sets received from the authors of three published studies presented data for the BDI score.^{38,47,124} The BDI evaluates key symptoms of depression including mood, pessimism, sense of failure, self-dissatisfaction, guilt, punishment, self-dislike, self-accusation, suicidal ideas, crying, irritability, social withdrawal, indecisiveness, body image change, work difficulty, insomnia, fatigability, loss of appetite, weight loss, somatic preoccupation and loss of libido.¹²⁵ The self-rated scale comprises 21 items, each of which is scored individually from 0 (least level of difficulty) to 3 (most level of difficulty). Scores are directly summed, with the total score ranging from 0 to 63 (higher scores indicating greater depressive severity). Grochtdreis *et al.* 2016 provide prediction models to map from BDI score to EQ-5D utility scores based on a sample of 1074 consecutive patients with depressive disorders from a psychotherapeutic outpatient clinic in Germany.¹²⁶ The authors estimated five prediction models with varying independent variables: BDI index (model 1), BDI index and age (model 2a), BDI index and grouped age (model 2b), BDI index and gender (model 3) and BDI index, gender and age (model 4). The authors reported that Models 2a and 4 showed the best predictive abilities (lowest root mean square errors). Therefore, model 4, with explanatory variables for BDI score, gender and age, was used to estimate EQ-5D utility scores with the following equation:

$$\text{EQ-5D score} = 1.0779 - 0.0161 \times \text{BDI score} - 0.0175 (\text{Males}) - 0.0027 \times \text{Age}$$

Data assessed at 28,¹²⁴ 30⁴⁷ and 35³⁸ weeks received from the collaborators for the three relevant published studies were grouped and analysed as a single time point ($N = 247$). A mixed-effect regression model (random effects on study and fixed effects on participants) was run to estimate a difference of 0.0395 (95% CI 0.013 to 0.046) in utility score between TRT and SoC. The regression predicted utility score was used for SoC group (0.766) and the difference added to estimate the TRT score. Utility multipliers were obtained by dividing the utility scores for TRT and SoC by the EQ-5D population norm for the regression sample age.¹²³ These multipliers were applied to the Ara and Brazier general population formula to obtain the age-specific utility scores applied to the TRT and SoC *No complications* health states.¹²³ Table 23 shows the utility scores by starting age used in the model.

TABLE 23 Utility scores for the *No Complication* health state for TRT and SoC

TRT	Coefficient	SE	95% CI	Distributional form
SF-6D	0.0042	0.0084	(-0.012 to 0.021)	Normal
Mapped EQ-5D	0.0295	0.0087	(0.013 to 0.046)	Normal
	TRT	SoC		
SF-6D based utility scores				
40 years old	0.910	0.905		
60 years old	0.838	0.834		
75 years old	0.767	0.763		
BDI based utility scores				
40 years old	0.854	0.823		
60 years old	0.787	0.758		
75 years old	0.720	0.693		

Further to the QoL data discussed above, five data sets presented data on the IIEF-15. While there is no mapping algorithm linking the IIEF-15 to the EQ-5D, the instrument has been used to develop health states that have been valued using the Time Trade-Off (TTO) method.¹²⁷ Stolk *et al.* (2000) conducted a cost-utility analysis of sildenafil compared with papaverine-phenolamine injections for the treatment of erectile dysfunction. The authors elicited 24 health states using two of the 15 questions from the International Index of Erectile Function (IIEF): the ability to attain an erection (question 3) and ability to maintain an erection (question 4). The use of these utility weights to value individuals' health states according to IIEF data was discussed by the members of the advisory group for this project (4th Advisory Group Meeting, 2 December 2019). It was agreed that the quality-of-life dimensions relevant for the population of interest were broader than the two questions considered in the TTO exercise conducted by Stolk *et al.* 2000. Therefore, it was felt that the use of the IIEF data to estimate utility scores was less relevant than the directly measured sources (SF-6D) or the values obtained from the mapping algorithm through the effect on depressive symptoms.

Health service resource use and costs

Testosterone replacement therapy

Resource use associated with medication, administration (when applicable) and monitoring were considered in the model. Oral capsules, gel applied on the skin or by intramuscular injection are available to administer TRT. The use of oral capsules is very limited in the UK.¹²⁸ Therefore, the model considered the four medicines most widely prescribed in the UK: Testogel[®], Tostran[®], Sustanon[®] and Nebido[®]. British National Formulary (BNF) and the NHS indicative prices were used to value these medicines.

Testogel[®] 16.2 mg/g gel [Besins Healthcare (UK) Ltd] is administered by a pump with one pump actuation delivering 1.25 g of gel containing 20.25 mg of testosterone.¹²⁹ Two pumps per day were assumed to deliver the daily doses (Dr Channa Jayasena, personal communication, 25 May 2020) with each prescription lasting 5 weeks at a cost of £31.11. Similarly, Tostran[®] 2% gel (Kyowa Kirin Ltd) is delivered with a canister piston, with one press of the canister piston delivering 0.5 g of gel containing 10 mg testosterone. Four presses were assumed to deliver the daily needed doses with each prescription, at a cost of £28.62, lasting 4 weeks.

Sustanon[®] 250 mg/ml solution injection (Aspen Pharma Trading Ltd) is administered once a month (12 injections in a year; Dr Channa Jayasena, personal communication, 25 May 2020) at a cost of £2.45 per injection. The model assumed that 50% of patients would self-administer and 50% would be delivered by a nurse. Nebido[®] 1000 mg/4 ml solution injection (Bayer Plc) costs £87.11 per injection and is always administered by a nurse or a doctor. A loading phase was assumed at start of treatment with one injection at the start, then at 6 weeks, and then every 3 months (4 per year) thereafter. For simplicity, all administrations were assumed to be delivered by a nurse. Therefore, the cost of TRT injection administration assumed 15 minutes of a nurse at a cost of £11.38 (i.e. 50% hospital nurse Band 6 at £49 per hour and 50% GP nurse at £42 per hour).

Finally, the annualised defined daily doses reported by Heald and colleagues were used to define the proportion of people using each type of administration in the model:¹²⁸ namely, 29%, 15%, 8% and 48%, for Testogel[®], Tostran[®], Sustanon[®] and Nebido[®], respectively.

Monitoring of individuals under TRT

Clinical guidelines were searched for an agreed monitoring and testing schedule for TRT. The information sheet for primary care prescribers for TRT for adult males with hypogonadism from the Nottinghamshire Area Prescribing Committee presents a schedule for monitoring and testing by mode of administration.¹³⁰ This schedule was discussed by the clinicians in the project management group, with the group agreeing the schedule was reasonable and could be used in the economic model. Therefore, for individuals using testosterone in gel a testosterone level (Tlevel) test was assumed at 4–6 weeks after the start of treatment (see [Table 5](#)). At 3–6 months the following physiological tests were

assumed: Tlevel, prostate-specific antigen (PSA; with a digital rectal examination if clinically indicated), Hb and haematocrit, liver function test (LFT) and lipid profile tests (LPTs). Further, all tests were assumed at 3–6 months and at 4 months for individuals using Sustanon and Nebido, respectively. For all products, all tests were assumed to be conducted at 12 months and annually thereafter. The unit costs for these tests were obtained from the National Schedule of NHS Costs 2019–20, Directly Accessed Pathology Services (see [Table 5](#)). A phlebotomy cost of £4.77 (DAPS08) was added for blood extraction.

Hypogonadism patients are monitored either at hospital as outpatients or at a primary care practice. It was assumed that half of patients would be monitored at the hospital and 50% in the community. For those individuals who were monitored as hospital outpatients, half of the hospital visits were assumed to happen at the endocrinology service and half at the urology service (see [Table 24](#)).

Standard of care

The standard treatment for individuals with hypogonadism is TRT. In the absence of TRT, it is unclear how often hypogonadal individuals would seek medical advice to deal with symptoms resulting from their clinical condition. As the main symptom for hypogonadism is a reduction in sexual function, one GP visit was assumed for the SoC strategy based on the advice of clinical experts in the project management group. Moreover, 96% of the SoC cohort were assumed to use medications for erectile

TABLE 24 Monitoring test schedule. Unit costs for tests and monitoring visits

Timing of monitoring	Tests to be done					
	Testosterone level	PSA (+DRE if clinically indicated) in men > 40 years	Hb and haematocrit	LFT	Lipid profile	Phlebotomy
Baseline (all products)	Yes	Yes	Yes	Yes	Yes	Yes
At 4–6 weeks (gel only)	Yes					Yes
At 3–6 months (gel or Sustanon only)	Yes	Yes	Yes	Yes	Yes	Yes
At 4 months (Nebido only, i.e. pre-3rd dose)	Yes	Yes	Yes	Yes	Yes	Yes
At 12 months (all products)	Yes	Yes	Yes	Yes	Yes	Yes
Annually thereafter (all products)	Yes	Yes	Yes	Yes	Yes	Yes
Unit costs	£1.22	£1.22	£2.58	£1.22	£1.22	£4.77
Currency code	DAPS04		DAPS05	DAPS04		DAPS08

Monitoring visits	Unit costs (£)	Notes, source:
GP visit	£39	GP – per surgery consultation lasting 9.22 minutes. PSSRU – Unit Costs of Health and Social Care 2020
Hospital visit – urology	£111	Total for Service code 101. National Schedule of NHS Costs Year: 2019–20 – all NHS trusts and NHS foundation trusts – Outpatient Attendances Data
Hospital visit – endocrinology	£162	Total for Service code 302. National Schedule of NHS Costs Year: 2019–20 – all NHS trusts and NHS foundation trusts – Outpatient Attendances Data

DRE, digital rectal examination; PSSRU, Personal and Social Services Unit. Source: National Schedule of NHS Costs 2019–20; Directly Accessed Pathology Services

dysfunction according to the findings by Rosen and colleagues when assessing the symptoms of hypogonadism using a self-administered questionnaire.⁸⁴ In addition, the numbers of items prescribed per 1000 men reported by Bell and colleagues were used to allocate, proportionately, the associated monthly cost for sildenafil (£1.27 for four tablets – one per week) and tadalafil (£4.66 per 28 tablets – one per day).^{131,132} The authors reported that 125.8 items per 1000 men of sildenafil were prescribed through the NHS in 2019. The items prescribed for tadalafil (29.3) were read from the authors' [Figure 1](#) using a webpage tool (www.graphreader.com).¹³³ A reduction of 63% of erectile dysfunction medications was assumed for individuals under TRT.¹³⁴

Health state utilities and costs associated with complications

The unit cost and utilities associated with CV, CBV and the peripheral vascular system events were sourced through searches of technology appraisals, clinical guidelines and health technology assessments on the NICE and National Institute for Health and Care Research (NIHR) websites. These sources were favoured as they are based on comprehensive literature reviews related to the condition of interest, utilise large data sets of UK patients and have been used in the NHS decision-making process. Following the method used in NICE Clinical Guidelines (CG181), each complication is attributed a short- and long-term cost and utility multiplier.¹³⁵ Hence, patients accrue alternative costs and utilities in the short and long term for each condition depending on whether it can be considered an ongoing or instantaneous health event.

Health state utility due to complications

The utility multipliers applied in the short and long term for each event are reported in [Table 25](#). Long-term conditions such as arrhythmia, CHD, valvular heart disease and peripheral vascular disease attribute

TABLE 25 Utility multipliers attached to CV complications

Variable	Point estimate	Standard error	Distributional form	Source
Arrhythmia ^a	0.81	0.081 ^b	Normal	NICE CG181
CHD ^a	0.84	0.002	Normal	NICE NG185
Heart failure ^a	0.68	0.02	Normal	NICE CG181
Myocardial infarction first year	0.76	0.018	Normal	NICE CG147
Myocardial infarction subsequent cycles	0.88	0.018	Normal	NICE CG147
Cardiac arrest	0.71	0.019	Normal	Perkins <i>et al.</i> 2021
New angina ^c and unstable angina first 6 months	0.77	0.038	Normal	NICE CG181
New angina ^c and unstable angina subsequent cycles	0.88	0.018	Normal	NICE CG181
Stable angina ^a	0.81	0.038	Normal	NICE CG181
Aortic aneurysm and aortic dissection ^d first 3 months	0.90	0.090 ^b	Normal	NICE NG156
Aortic aneurysm and aortic dissection ^d subsequent cycles up to 1 year	0.95	0.095 ^b	Normal	NICE NG156
Valvular heart disease ^a	0.81	0.081 ^b	Normal	DG14
Peripheral vascular disease ^a	0.81	0.038	Normal	NICE CG181
Transient ischaemic attack ^a	0.9	0.03	Normal	NICE CG181
Ischaemic stroke ^a	0.63	0.04	Normal	NICE CG181

a Same multiplier used for event and subsequent cycles.

b Standard error of 10% assumed where it was not reported.

c New angina assumed equal outcomes as unstable angina.

d Aortic dissection assumed equal outcomes as aortic aneurysm.

equal short- and long-term utility multipliers. The majority of these multipliers were sourced from NICE Clinical Guidelines on CV disease (CG181), acute coronary syndromes (NG185), the diagnosis and management of peripheral arterial disease (CG147), abdominal aortic aneurysm (NG156) and atrial fibrillation and heart valve disease (DG14).^{117,135-138} It has been assumed that cardiac arrests occur out of hospital due to its sudden onset. Therefore, the multiplier for cardiac arrest was sourced from the PARAMEDIC2 RCT.¹³⁹ This trial recruited 8014 participants from across England and Wales with an out-of-hospital cardiac arrest from 2014 to 2017. The primary outcome was the rate of survival at 30 days and secondary outcomes included the rate of survival until hospital discharge. For simplicity, the long-term utility multiplier was used from the moment of the event for individuals surviving a cardiac arrest, as these patients would be clinically dead (utility score = 0) for a very short period of time only.

Costs of complications

The instantaneous event cost and annualised long-term cost of complications are presented in [Table 26](#). These were determined using UK clinical guidelines, NHS reference costs, the BNF, and the Personal and Social Services Unit (PSSRU).^{132,140,141} All costs are expressed in 2019/2020 prices using the PSSRU inflation indices.¹⁴¹

Cardiovascular and cerebrovascular events

The most common form of arrhythmia is atrial fibrillation,¹⁴² therefore, the cost of atrial fibrillation was used as the unit cost of arrhythmia. The unit cost was sourced from a large UK study which utilised data from the Information Services Division on GP visits, outpatient visits and hospitalisation rates in patients with atrial fibrillation to calculate the average direct cost per patient to the NHS.¹⁴³ CHD has been assumed to be stable where the annual cost is equal from diagnosis. This cost was sourced from a large cohort study of 94,966 patients in England from 2001 to 2010 using the Clinical Practice Research Database (CPRD) and Hospital Episode Statistics (HES).¹⁴⁴ NHS reference costs, PSSRU and the Prescription Cost Analysis (PCA) were used to estimate the direct annual cost of CHD to the NHS. The costs of heart failure, myocardial infarction and CBV events were sourced from Danese *et al.* 2016.¹⁴⁵ This is a large UK retrospective cohort study of 24,093 patients who experienced their first CV event from January 2005 to March 2012. The study sourced data from the CPRD and the HES to estimate the direct cost of treatment to the NHS.

Like the utility multiplier attributed to cardiac arrest, the unit cost was sourced from the PARAMEDIC2 RCT. Given that the cardiac arrests reported in [Table 21](#) did not result in death, the cost applied incorporates all healthcare resource use utilisation for patients who survive to hospital discharge up to 6 months. Furthermore, cardiac arrest patients do not incur further health service costs, as it is assumed to be an instantaneous health event.

The health economic appendix of CG181 for CV disease reports the estimated 6-month and 1-year post-event cost of stable and unstable angina.¹³⁵ These are based upon NHS standard practice with the PSSRU, Healthcare Resource Group (HRG) and BNF used as the components of the unit costs. It has been assumed that 'new angina' incurs the same resource use as 'unstable angina' as symptoms are sudden onset as is the definition of unstable angina.

Events of the peripheral vascular system

In the absence of explicit unit costs reported in the clinical guidelines for diseases of the peripheral vascular system, unit costs were constructed utilising the NICE recommendations of care and the relevant HRG and PSSRU costs. For simplicity, valvular heart disease is assumed to be aortic stenosis, which is one of the two most common forms of the disease in the UK.¹⁴⁶ As advised in CG187 for acute heart failure, aortic stenosis is treated with heart valve replacement surgery.¹⁴⁷ Therefore, the unit cost in the short and long term is composed of the HRG cost of surgery and cardiology follow-up appointments, respectively. Similarly, aortic dissection and aneurysm are recommended in the NICE Guideline 156 to be treated using elective open surgical repair unless contraindicated by other

TABLE 26 Unit costs attached to CV complications

Type of complication	Short-term unit cost (£) Mean (SE) ^a	Annualised long-term unit cost (£) Mean (SE) ^a	Source
Arrhythmia	N/A	£1322 (132)	Burdett and Lip, 2020 ¹⁴³
CHD	N/A	£2002 (4145)	Walker <i>et al.</i> 2016 ¹⁴⁴
Heart failure ^b	£4324 (121)	£3911 (340)	Danese <i>et al.</i> 2016 ¹⁴⁵
Myocardial infarction ^b	£5743 (125)	£2548 (180)	Danese <i>et al.</i> 2016 ¹⁴⁵
Cardiac arrest ^b	£35,317 (7716)	N/A	Perkins <i>et al.</i> 2021 ¹³⁹
New angina ^{b,c}	£3664 (366)	£426 (44)	CG181 Health economic appendix ¹³⁵
Unstable angina ^b	£3664 (366)	£426 (44)	CG181 Health economic appendix ¹³⁵
Stable angina ^b	£8555 (856)	£265 (27)	CG181 Health economic appendix ¹³⁵
Aortic aneurysm ^d	£12,834 (1283)	£262 (26)	NG156 Health economic appendix, Cardiac follow-up F2F, NHS reference costs 2019/20 ^{137,148}
Aortic dissection ^{d,e}	£12,834 (1283)	£262 (26)	NG156 Health economic appendix, Cardiac follow-up F2F, NHS reference costs 2019/20 ^{137,148}
Valvular heart disease ^d	£12,874 (1287)	£262 (26)	Weighted average of ED25, Cardiac follow-up F2F, NHS reference costs 2019/20 ¹⁴⁸
Peripheral vascular disease ^f	£516 (52)	£185 (19)	Combination of CG147 Health economic appendix, PSSRU GP visit cost, BNF cost of year supply of atorvastatin and aspirin ^{132,141,149}
Transient ischaemic attack ^b	£2698 (97)	£2561 (166)	Danese <i>et al.</i> 2016 ¹⁴⁵
Ischaemic stroke ^b	£4758 (111)	£2739 (300)	Danese <i>et al.</i> 2016 ¹⁴⁵

a Standard error of 10% assumed where it was not reported.

b Short-term cost for 6 months.

c New angina assumed equal cost as unstable angina.

d Short-term cost for 2 months.

e Aortic dissection assumed equal cost as aortic aneurysm.

f Short-term cost for 12 months.

Note

All costs reported in 2020 prices, uplifted by the PSSRU inflation indices where relevant. Gamma distributions based on mean and standard error were defined for all unit costs.

comorbidities.¹³⁷ Therefore, the cost of surgery reported in the NICE Guideline 156 was attributed as the short-term cost, whereas the long-term cost consisted of the recommended follow-up appointments for long-term care. Peripheral vascular disease, described as peripheral arterial disease in the NICE Clinical Guideline 147, is commonly characterised by a painful condition in the legs called intermittent claudication.¹³⁶ The treatment for this is a supervised or unsupervised exercise regime, statins and blood thinners. In the long term, it was assumed that patients incur the cost of the supervised exercise regime reported in CG147, the cost of medication and the cost of the GP follow-up visits.

Model validation

Several steps were taken in order to secure the quality of the model.¹⁵⁰ The model structure was agreed with the members of the Advisory Groups for this project to secure the model structure face validity. White-box testing materialised throughout the whole model implementation (e.g. verification of formulae results with an external software) and black-box verification tests were conducted to assess the behaviour of the model specific input values (e.g. all utility scores equal to 1 to check whether

total QALYs equal total life years). Markov traces were extracted from the modelling software and the cumulative proportion entering the complication states (i.e. *Cardiac pathology*, *Pathology of the peripheral vascular system* and the *Pathology of the cerebrovascular system* Markov states) were added up. These cumulative proportions at 10 years (120 cycles) were compared against the 10-year risk of experiencing a first CV event obtained from the QRisk3 CV risk calculator.¹⁵¹ The model showed a cumulative proportion of 5%, 15.9% and 29.8% entered the model complication states at 10 years for the 40-, 60- and 75-year-old cohort, respectively. These figures were compared with a weighted average of the QRisk3 risk calculator score obtained for a white male with diagnosis of erectile dysfunction (73%) and a white man with diagnosis of erectile dysfunction and diabetes (27%). The model risk figures are higher than the weighted averages (i.e. 3.1%, 13.6% and 27.2% for 40-, 60- and 75-year-olds, respectively). However, the TestES model population included other comorbidities such as smoking resulting in a higher risk of CV and CBV events. The model data for the risk of CV and CBV event complications were sourced from the British Heart Foundation and were the latest data available, therefore no calibration was conducted.

Model analysis

The analysis captures the cumulative health and social care costs from the perspective of the NHS and QALYs accrued by patients under TRT and SoC strategies over a 10-year time horizon.

The model was run probabilistically to characterise the joint uncertainty in the modelled outputs (cost and QALYs) arising from the uncertainty in the input parameters. Monte Carlo simulation techniques were used to analyse the model many times (10,000 iterations), with sets of values drawn at random from the probability distributions assigned to each model parameter. To characterise the uncertainty in the mean parameter values, beta, gamma and log-normal distributions were used for probabilities, costs and relative effects, respectively. Normal distributions were used for the TRT utility difference estimated from the regression analysis of the TestES data and the utility multipliers attached to complications. Details of the probability distributions used are reported within [Tables 20–26](#). Ten thousand iterations were deemed enough to obtain stable results. The outputs from these probabilistic analyses are presented as the probability of TRT and SoC being cost-effective at £10,000, £20,000 and £30,000 cost-effectiveness threshold values. Cost-effectiveness acceptability curves (CEACs) were also produced for selected scenarios to further illustrate the uncertainty around the model results.¹⁵² CEACs present the probability of the compared strategies generating the greatest net monetary benefit for different cost-effectiveness thresholds (cost per QALY gained). In addition, incremental cost-effectiveness ratios (ICERs) were estimated to compare TRT against SoC. The ICER is defined as the ratio of the difference in expected costs over the difference in expected QALYs between TRT and SoC. As the model was run probabilistically, cost and QALYs were averaged across the 10,000 iterations and the ICER was calculated with the average difference in costs and QALYs between TRT and SoC.

Key assumptions

Results are reported for eight scenarios defined according to alternative assumptions around key model effectiveness parameters for TRT versus SoC: the definition of the utility difference between TRT and SoC, the RR of mortality and RR of CV, peripheral vascular and CBV complications. In these scenarios, the length of time at which the different effect estimates were applied varied: 12 months, 10 years, non-existent or patient lifetime. Results are reported for three age groups: 40-, 60- and 75-year-olds.

Details for the eight scenarios are defined below and summarised in [Table 27](#):

- Scenario 1: the utility score difference between TRT and SoC was based on the SF-6D data derived from the analysis of TestES IPD. Moreover, the RR of mortality (RR mortality) was defined as 0.46 (95% CI 0.18 to 1.25) and the RR of experiencing a complication (RR complication) equal to 1.06 (95% CI 0.82 to 1.38). The model time horizon was defined as 10 years; however, all these effects were applied for the first 12 months only to be consistent with the clinical effectiveness primary outcomes.

TABLE 27 Summary of the economic evaluation scenarios

Scenario	Utility difference	RR mortality	RR complications	Time period for effects	Model time horizon
1	SF-6D	0.46	1.06	1 year	10 years
2a	SF-6D	0.46	1.06	10 years	10 years
2b	BDI mapped to EQ-5D	0.46	1.06	10 years	10 years
3a	SF-6D	1	1	10 years	10 years
3b	BDI mapped to EQ-5D	1	1	10 years	10 years
4	No difference	0.46	1	10 years	10 years
5	No difference	1	1.06	10 years	10 years
6a	SF-6D	1	1.06	1 year	10 years
6b	BDI mapped to EQ-5D	1	1.06	1 year	10 years
7a	SF-6D	1	1.06	10 years	10 years
7b	BDI mapped to EQ-5D	1	1.06	10 years	10 years
8	SF-6D	1	1.06	Lifetime	Lifetime

- Scenario 2a: similar definitions for utility difference, RR mortality and RR complication as for scenario 1 were used. However, these effects were maintained for the whole 10-year time horizon.
- Scenario 2b: all definitions as for Scenario 2a except for the utility difference between TRT and SoC that was based on the BDI score mapped to the EQ-5D score.
- Scenario 3a: the utility score difference based on SF-6D was applied for 10 years. No difference in mortality or complication rates was considered (i.e. RR mortality equal to one and RR complication equal to one).
- Scenario 3b: all definitions as for scenario 3a except for the utility difference between TRT and SoC that was based on the BDI score mapped to the EQ-5D score.
- Scenario 4: the RR for mortality (0.46) was applied for 10 years. No utility score difference and no difference in complication rates (RR complication equal to one).
- Scenario 5: the RR for complications (1.06) was applied for 10 years; no utility score difference and no difference in mortality rates (RR mortality equal to one).
- Scenario 6a: the utility score difference based on the SF-6D and RR for complications (1.06) was applied for 1 year. No difference in mortality (i.e. RR mortality equal to one).
- Scenario 6b: all definitions as for Scenario 6a except for the utility difference between TRT and SoC that was based on the BDI score mapped to the EQ-5D score.
- Scenario 7a: the utility score difference based on SF-6D and the RR for complications (1.06) was applied for 10 years. No difference in mortality (i.e. RR mortality equal to one).
- Scenario 7b: all definitions as for Scenario 7a except for the utility difference between TRT and SoC that was based on the BDI score mapped to the EQ-5D score.
- Scenario 8: lifetime QoL and rate of complication differences were assumed. The SF-6D utility score difference was used as this was the smaller observed difference between TRT and SoC. No difference in mortality (i.e. RR mortality equal to one).

Results

Results for the eight scenarios for the 40-, 60- and 75-year-old cohorts are reported in [Tables 28–30](#), respectively. Average cost, incremental cost, average QALY, incremental QALYs and ICERs are presented together with the probability of the strategies being cost effective at £10,000, £20,000 and £30,000 per QALY gained thresholds.

TABLE 28 Scenario analysis results. Model starting age 40 years

Strategy	Cost (£)	Incr. cost (£)	QALYs	Incr. QALYs	ICER (£)	Probability of cost-effective at alternative cost-effectiveness thresholds (%)		
						10,000	20,000	30,000
Scenario 1: 1-year SF-6D utility score difference, RR mortality = 0.46 and RR complication = 1.06								
SoC	922		7.415			100.0	100.0	100.0
TRT	4158	3236	7.424	0.009	357,797	0.0	0.0	0.0
Scenario 2a: as scenario 1 but differences applied for 10 years								
SoC	923		7.415			100.0	90.0	72.0
TRT	4192	3269	7.478	0.063	51,674	0.0	10.0	28.0
Scenario 2b: as scenario 2a but using BDI-based utility score difference								
SoC	922		6.739			69.3	7.2	1.7
TRT	4190	3268	7.024	0.285	11,479	30.7	92.8	98.3
Scenario 3a: 10-year SF-6D utility score difference; no difference in mortality or complication (RRs = 1)								
SoC	922		7.415			100.0	95.2	82.9
TRT	4151	3229	7.449	0.034	95,005	0.0	4.8	17.1
Scenario 3b: as scenario 3a but using BDI-based utility score difference								
SoC	922		6.739			78.8	10.5	2.7
TRT	4151	3229	6.999	0.260	12,429	21.2	89.5	97.3
Scenario 4: RR for mortality (0.46) applied for 10 years; no utility score difference; RR for complication = 1								
SoC	922		7.415			100.0	100.0	100.0
TRT	4168	3246	7.447	0.032	100,946	0.0	0.0	0.0
Scenario 5: RR for complications (1.06) applied for 10 years; no utility score difference; RR for mortality = 1								
SoC	921		7.415			100.0	100.0	100.0
TRT	4172	3251	7.411	-0.003	-1,012,178	0.0	0.0	0.0

Strategy	Cost (£)	Incr. cost (£)	QALYs	Incr. QALYs	ICER (£)	Probability of cost-effective at alternative cost-effectiveness thresholds (%)		
						10,000	20,000	30,000
Scenario 6a: 1-year SF-6D utility score difference and RR complications = 1.06. RR for mortality = 1								
SoC	922		7.415			100.0	100.0	100.0
TRT	4155	3233	7.418	0.003	927,285	0.0	0.0	0.0
Scenario 6b: as scenario 6a but using BDI-based utility score difference								
SoC	923		6.739			100.0	100.0	100.0
TRT	4156	3233	6.771	0.032	102,463	0.0	0.0	0.0
Scenario 7a: 10-year SF-6D utility score difference and RR complications = 1.06. RR for mortality = 1								
SoC	924		7.415			100.0	96.1	84.5
TRT	4175	3251	7.445	0.030	108,362	0.0	3.9	15.5
Scenario 7b: as scenario 7a but using BDI-based utility score difference								
SoC	922		6.739			81.2	11.8	3.2
TRT	4173	3252	6.994	0.255	12,741	18.8	88.2	96.8
Scenario 8: SF-6D utility score difference and RR complications = 1.06, applied for lifetime. RR for mortality = 1. Lifetime time horizon.								
SoC	4633		16.981			99.8	91.5	80.1
TRT	10,842	6210	17.013	0.032	192,861	0.2	8.6	19.9
Incr., Incremental.								

TABLE 29 Scenario analysis results. Model starting age 60 years

Strategy	Cost (£)	Incr. cost (£)	QALYs	Incr. QALYs	ICER (£)	Probability of cost-effective at alternative cost-effectiveness thresholds (%)		
						10,000	20,000	30,000
<i>Scenario 1: 1-year SF-6D utility score difference, RR mortality = 0.46 and RR complication = 1.06</i>								
SoC	1573		6.424			100.0	100.0	100.0
TRT	4580	3007	6.450	0.026	113,992	0.0	0.0	0.0
<i>Scenario 2a: as scenario 1 but differences applied for 10 years</i>								
SoC	1575		6.424			94.1	44.6	26.7
TRT	4710	3135	6.585	0.161	19,444	6.0	55.4	73.3
<i>Scenario 2b: as scenario 2a but using BDI-based utility score difference</i>								
SoC	1578		5.839			34.0	4.3	2.0
TRT	4712	3134	6.187	0.348	8993	66.0	95.8	98.0
<i>Scenario 3a: 10-year SF-6D utility score difference; no difference in mortality or complication (RRs = 1)</i>								
SoC	1569		6.424			100.0	96.3	85.3
TRT	4551	2982	6.454	0.030	100,625	0.0	3.8	14.7
<i>Scenario 3b: as scenario 3a but using BDI-based utility score difference</i>								
SoC	1573		5.838			85.9	13.2	3.2
TRT	4555	2982	6.063	0.225	13,258	14.1	86.8	96.8
<i>Scenario 4: RR for mortality (0.46) applied for 10 years; no utility score difference; RR for complication = 1</i>								
SoC	1572		6.424			100.0	47.7	24.6
TRT	4651	3079	6.565	0.141	21,845	0.0	52.3	75.4
<i>Scenario 5: RR for complications (1.06) applied for 10 years; no utility score difference; RR for mortality = 1</i>								
SoC	1577		6.424			100.0	100.0	100.0
TRT	4615	3039	6.416	-0.008	-359,489	0.0	0.0	0.0

Strategy	Cost (£)	Incr. cost (£)	QALYs	Incr. QALYs	ICER (£)	Probability of cost-effective at alternative cost-effectiveness thresholds (%)		
						10,000	20,000	30,000
Scenario 6a: 1-year SF-6D utility score difference and RR complications = 1.06. RR for mortality = 1								
SoC	1571		6.424			100.0	100.0	100.0
TRT	4561	2990	6.426	0.002	1,327,419	0.0	0.0	0.0
Scenario 6b: as scenario 6a but using BDI-based utility score difference								
SoC	1566		5.838			100.0	100.0	100.0
TRT	4556	2990	5.867	0.028	105,804	0.0	0.0	0.0
Scenario 7a: 10-year SF-6D utility score difference and RR complications = 1.06. RR for mortality = 1								
SoC	1572		6.424			100.0	96.9	87.8
TRT	4610	3038	6.446	0.022	136,382	0.0	3.2	12.2
Scenario 7b: as scenario 7a but using BDI-based utility score difference								
SoC	1574		5.838			87.7	18.5	5.1
TRT	4610	3036	6.055	0.217	14,008	12.4	81.5	94.9
Scenario 8: SF-6D utility score difference and RR complications = 1.06, applied for lifetime. RR for mortality = 1. Lifetime time horizon.								
SoC	4326		10.707			99.7	92.1	82.7
TRT	8897	4571	10.712	0.005	913,465	0.4	7.9	17.3
Incr., incremental.								

TABLE 30 Scenario analysis results. Model starting age 75 years

Strategy	Cost (£)	Incr. cost (£)	QALYs	Incr. QALYs	ICER (£)	Probability of cost-effective at alternative cost-effectiveness thresholds (%)		
						10,000	20,000	30,000
<i>Scenario 1: 1-year SF-6D utility score difference, RR mortality = 0.46 and RR complication = 1.06</i>								
SoC	2415		4.738			100.0	89.3	49.2
TRT	4810	2395	4.810	0.072	33,105	0.0	10.7	50.8
<i>Scenario 2a: as scenario 1 but differences applied for 10 years</i>								
SoC	2410		4.738			24.9	12.7	10.2
TRT	5188	2778	5.148	0.410	6778	75.1	87.3	89.8
<i>Scenario 2b: as scenario 2a but using BDI-based utility score difference</i>								
SoC	2404		4.306			11.6	5.7	4.6
TRT	5182	2778	4.841	0.534	5198	88.4	94.3	95.4
<i>Scenario 3a: 10-year SF-6D utility score difference; no difference in mortality or complication (RRs = 1)</i>								
SoC	2413		4.737			100.0	97.5	87.8
TRT	4719	2306	4.759	0.022	106,983	0.0	2.6	12.2
<i>Scenario 3b: as scenario 3a but using BDI-based utility score difference</i>								
SoC	2411		4.306			91.1	16.4	4.3
TRT	4717	2306	4.470	0.165	14,010	8.9	83.6	95.8
<i>Scenario 4: RR for mortality (0.46) applied for 10 years; no utility score difference; RR for complication = 1</i>								
SoC	2410		4.738			21.6	11.3	9.4
TRT	5099	2689	5.149	0.412	6532	78.4	88.7	90.6
<i>Scenario 5: RR for complications (1.06) applied for 10 years; no utility score difference; RR for mortality = 1</i>								
SoC	2410		4.738			100.0	99.9	99.0
TRT	4800	2390	4.718	-0.020	-121,604	0.0	0.1	1.0

Strategy	Cost (£)	Incr. cost (£)	QALYs	Incr. QALYs	ICER (£)	Probability of cost-effective at alternative cost-effectiveness thresholds (%)		
						10,000	20,000	30,000
Scenario 6a: 1-year SF-6D utility score difference and RR complications = 1.06. RR for mortality = 1								
SoC	2411		4.738			100.0	100.0	100.0
TRT	4735	2324	4.736	-0.002	-1,216,873	0.0	0.0	0.0
Scenario 6b: as scenario 6a but using BDI-based utility score difference								
SoC	2417		4.306			100.0	100.0	100.0
TRT	4742	2325	4.327	0.021	109,280	0.0	0.0	0.0
Scenario 7a: 10-year SF-6D utility score difference and RR complications = 1.06. RR for mortality = 1								
SoC	2405		4.738			100.0	95.4	87.3
TRT	4795	2390	4.740	0.002	976,583	0.1	4.6	12.7
Scenario 7b: as scenario 7a but using BDI-based utility score difference								
SoC	2410		4.306			89.2	34.8	16.1
TRT	4804	2394	4.452	0.147	16,338	10.8	65.2	83.9
Scenario 8: SF-6D utility score difference and RR complications = 1.06, applied for lifetime. RR for mortality = 1. Lifetime time horizon.								
SoC	3085		5.655			99.6	92.3	84.1
TRT	5866	2782	5.641	-0.015	-190,434	0.4	7.7	15.9
Incr., incremental.								

Modelling for 40-year-old men with MH

The discounted average costs for scenario 1 are £1126 and £4215 for SoC and TRT, respectively, giving an incremental cost of £3089. The average costs are very similar for scenarios 1–7 where a 10-year time horizon was assumed. The Monte Carlo simulation seed number was not fixed for each model run; therefore, the small variation in the SoC average costs for scenarios 1–7 is explained by sampling variation. Discounted average QALYs for scenario 1 are 7.414 and 7.424 for SoC and TRT, respectively, resulting in a small QALY gain of 0.010 from TRT. The ICER for scenarios 1 is well above the usual cost-effectiveness thresholds used by government decision-making bodies such as NICE in the UK (i.e. £20,000 per QALY gained).¹¹⁴ The probability of TRT being cost-effective is zero at any threshold value because of the short term assumed for differences in the key parameters.

The differences in QoL, mortality and rate of complications were maintained for 10 years for scenario 2. While incremental cost and QALYs increased for this scenario, the increase in QALY is proportionally larger, reducing the ICERs for this scenario compared with scenario 1. However, the probability for TRT being cost-effective at £20,000 remains low (14%). BDI-based utilities were utilised in scenario 2b. This resulted in larger utility differences and greater QALY increments between TRT and SoC. Therefore, an ICER of £10,878 is obtained and the probability of TRT being cost-effective rises to 95% at £20,000 cost-effectiveness threshold. The QALY increment, ICER and the probability of being cost-effective are consistent for all the scenarios that assumed BDI-based utilities lasting 10 years (scenarios 2b, 3b, 7b).

The isolated impact of the difference in utility scores was considered in scenario 3. Cost and QALY increments reduced slightly compared with scenario 2. The ICER increased substantially for scenario 3a but not for scenario 3b; however, the probability of TRT being cost effective remains low for 3a and high for 3b, showing the limited impact that the differences in mortality and/or complications have for the cost-effectiveness of TRT in this age group. This is further illustrated in scenarios 4 and 5, where only mortality (scenario 4) or complication (scenario 5) differences were utilised. Both scenarios result in a zero probability of TRT being cost-effective at any threshold value.

A small number of deaths were reported within the TestES data set and the difference in mortality, upon discussion within the project team, was deemed as unreliable. Scenarios 6–8 show the impact of removing the mortality difference in the model. One- and 10-year QoL and rate of complication differences were assumed for scenarios 6 and 7, respectively. Results show an ICER below £20,000 only when the BDI-based utilities were assumed to last 10 years (scenario 7b). Finally, scenario 8 illustrates the effects of assuming a lifetime time horizon, and lifetime utility (SF-6D) and complication rate differences between TRT and SoC. Average cost and QALYs increased due to the longer time horizon considered. While the difference in cost doubled with respect to the shorter time horizon (scenario 7a), the difference in QALYs did not increase proportionally, resulting in a higher ICER (i.e. £119,436 compared with £91,580) and a probability of TRT being cost-effective of 11% at a threshold of £20,000.

Modelling for 60-year-old men with hypogonadism

Results for the 60-year-old group are reported in [Table 10](#). Discounted average costs for this cohort are higher than those reported in [Table 28](#) for the 40-year-old group due to the higher proportion of CV and CBV complications experienced by the older group. However, cost differences between TRT and SoC reduced slightly as the cohort ages. All-cause mortality increases for older groups, and therefore the 60-year-old cohort accrue relatively fewer QALYs. Crucially, as the underlying mortality for these age groups is higher, TRT appeared relatively more cost-effective for the scenarios that considered a reduction in mortality (scenarios 1, 2a, 2b and 4). Moreover, using BDI-based utilities (scenarios 2b, 3b, 6b and 7b) resulted in lower ICERs with respect to the corresponding scenarios using SF-6D utilities (scenarios 2a, 3a, 6a and 7a) and ICERs below the £20,000 threshold are obtained for the scenarios assuming long-term BDI-based utility differences (scenarios 2b, 3b and 7b). Furthermore, when the mortality difference was removed, the long-term BDI-based utility difference becomes crucial to cause the ICER to fall below the £20,000 threshold (scenario 7b). Finally, running the model for a lifetime time horizon did not improve the cost-effectiveness of TRT (scenario 8), as the cumulative difference in the

incidence of complications starts to outweigh any ongoing gains in general health state utility associated with TRT use.

Modelling for 75-year-old men with MH

In [Table 30](#), the results for the 75-year-old groups are presented. Discounted average costs are higher than those reported in [Tables 9](#) and [10](#) for the 40- and 60-year-old groups, respectively. Again, this is due to the higher proportion of CV and CBV complications experienced by the 75-year-old group. However, while average costs are higher for SoC and TRT, the cost differences are reduced further compared with the younger cohorts. Moreover, the 75-year-old cohort accrues fewer QALYs as the all-cause mortality further increases for the older group. Notably, the ICER for TRT falls below £20,000 for scenarios that considered a reduction in mortality (scenarios 1, 2a, 2b and 4) as the underlying mortality for this age group further increases, and the absolute benefit associated with relative mortality reduction increases. As for the younger cohorts, the ICERs decreased when BDI-based utilities were defined (scenarios 2b, 3b, 6b and 7b) compared with the corresponding scenarios using SF-6D utilities (scenarios 2a, 3a, 6a and 7a) and they fell below the £20,000 threshold for those scenarios that assumed a long-term BDI-based utility difference (scenarios 2b, 3b and 7b). As with the younger cohorts, long-term BDI based utilities are vital for the ICERs to fall below the £20,000 threshold when the mortality difference is removed, (scenario 7b). Lastly, a lifetime time horizon model scenario (scenario 8) showed that TRT was not cost-effective when applying SF-6D utilities.

Cost-effectiveness acceptability curves

The CEACs for TRT for the alternative scenarios for the 40-, 60- and 75-year-old starting age cohorts are reported in [Figures 18–20](#), respectively. The CEACs illustrate the decision uncertainty due to second-order uncertainty around the input parameter values. The curves show the probability of TRT being cost effective for a range of cost-effectiveness threshold values. The probability of TRT being cost-effective rises as the cost-effectiveness threshold increases, indicating that higher QALYs are obtained with TRT for almost all scenarios and age groups (see [Figures 18–20](#)). The exception is scenario 6, where only the difference in complications rates was assumed, which resulted in a greater number of TRT complications that were not overturned by a reduced mortality or a higher QoL.

The CEACs for scenarios where BDI-based utilities were used were drawn with short-dashed lines. For the 40-year-old cohort (see [Figure 18](#)), these CEACs showed a high probability of TRT being

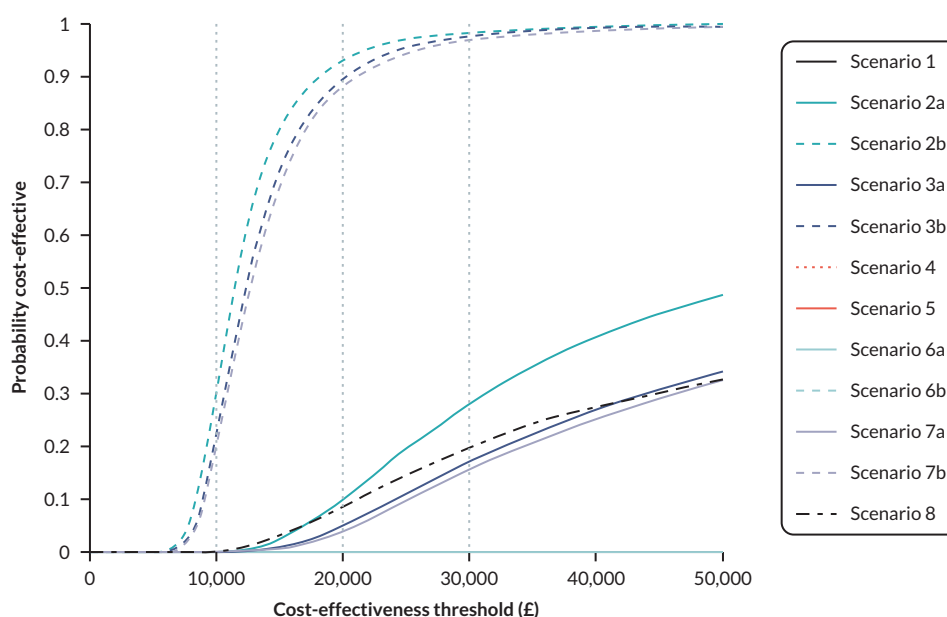


FIGURE 18 Cost-effectiveness acceptability curves for TRT for alternative scenarios – 40-year-old cohort.

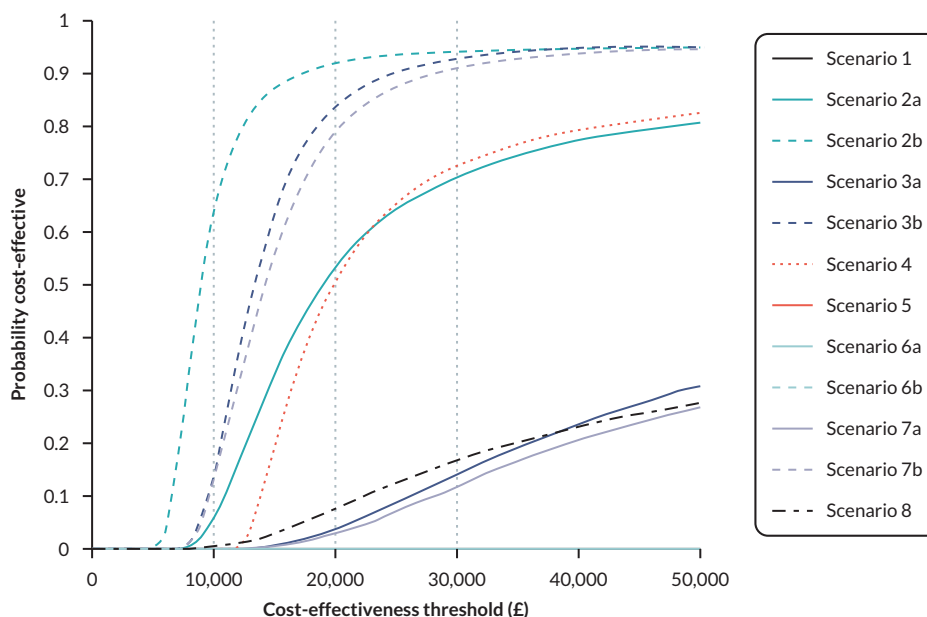


FIGURE 19 Cost-effectiveness acceptability curves for TRT for alternative scenarios –60-year-old cohort.

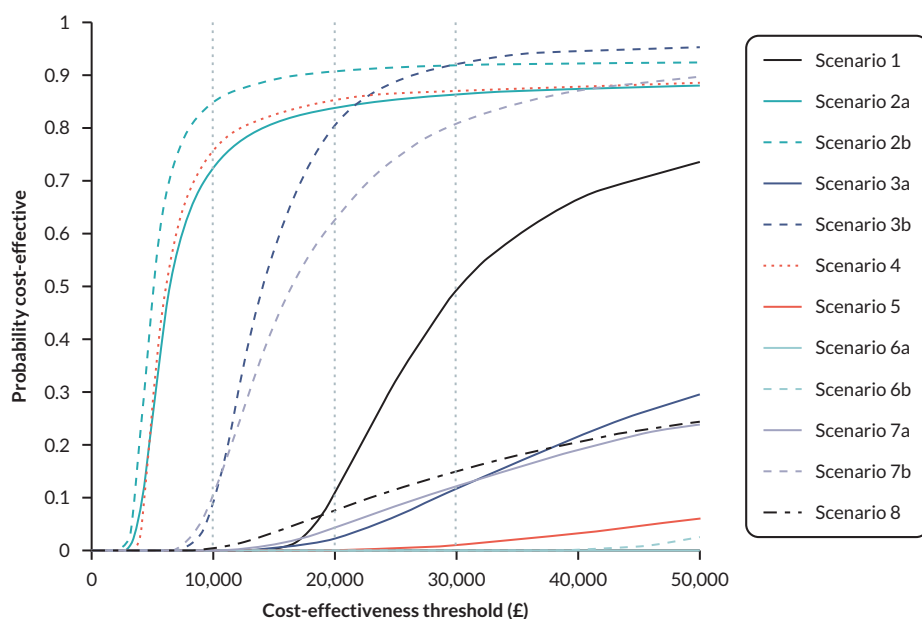


FIGURE 20 Cost-effectiveness acceptability curves for TRT for alternative scenarios – 75-year-old cohort.

cost-effective at a £20,000 threshold when the BDI-based utilities were applied for 10 years (scenarios 2b, 3b and 7b). The CEACs for the 60-year-old cohort group are presented in [Figure 19](#). Similarly, those strategies that used BDI based utilities for 10 years showed a high probability of TRT being cost-effective at a £20,000 threshold. In addition, the scenarios assuming a RR for mortality of 0.46 for 10 years (2a and 4) resulted in an over 50% probability of TRT being cost-effective at a £20,000 threshold. A low probability of cost-effectiveness was observed for TRT for those scenarios where SF-6D and no difference in mortality were defined (i.e. scenarios 3a, 6a and 7a). Furthermore, the CEACs for the 75-year-old cohort (see [Figure 20](#)) showed a high probability of TRT being cost-effective for those scenarios where BDI utility and/or relative mortality differences were assumed to last 10 years (scenarios 2a, 2b, 3b, 4 and 7b). Lastly, a lifetime time horizon together with lifetime relative effect on the complication rate and a lifetime increment in SF-6D utility were defined for scenario 8. This

scenario showed a very low probability for TRT being cost-effective regardless of the starting age of the population cohort (see [Figures 18–20](#)).

Discussion

The present chapter reports on the cost-effectiveness of TRT compared with no TRT. We conducted a cost-utility analysis from the UK NHS perspective. The Markov model developed incorporated the primary outcomes considered for the synthesis of clinical effectiveness evidence (see [Chapter 2](#)) and the utility scores obtained from the IPD analysis. A strength of the analysis is that key treatment effect parameters for the all-cause mortality, rate of CV and CBV complications and health state utility scores were obtained from the IPD analyses of the TestES data sets received from the trials investigators. To date, this is the largest RCT-based data set for men with hypogonadism comparing TRT with no TRT treatment.

Results show that the cost-effectiveness of TRT is dependent on the RR of all-cause mortality and the methods used to derive health state utility scores (i.e. through the SF-6D algorithm or a mapping exercise between BDI and EQ-5D scores) for the TRT versus SoC. When the RR of mortality favouring TRT and the BDI-based utility scores were used for the 10-year time horizon, ICERs fell below the £20,000 threshold, irrespective of the cohort starting age. ICERs were also below the £20,000 threshold for the 60-year-old and 75-year-old cohorts when the RR of mortality favouring TRT and the SF-6D utility were used with the differences lasting 10 years (scenario 2a). However, the ICER increased above this threshold when the difference in all-cause mortality between TRT and SoC was dropped, and the utility scores were defined using the SF-6D (scenario 7a). Extending the model time horizon for lifetime for the later scenario further increased the ICER, as the impact of complications became more pronounced, eroding the modest SF-6D-based QALY increment.

As mentioned above, the all-cause mortality RR was shown to be a driver of results, particularly for the older cohorts (60- and 75-year-old cohorts). This is due to the higher underlying mortality risk of these populations compared with that of the 40-year-old cohort. It is worth noting that this RR was based on a small number of events (i.e. 6 over 1621, 0.4% and 12 over 1537, 0.8%, for TRT and placebo groups, respectively); as such the results for the scenarios where a reduced risk of mortality for TRT was considered should be taken with extreme caution.

The RR of CV and CBV complications was also obtained from the TestES IPD meta-analysis. The estimate for the RR of complications was based on events in 120 individuals over 1601 (7.5%) and 110 individuals over 1519 (7.2%) for the TRT and placebo groups, respectively. This result was not statistically significant and there was no evidence to support a meaningful difference between TRT and SoC. However, the best estimate for the RR of complications was applied in the economic model in the context of a probabilistic analysis that appropriately characterised the uncertainty around the point estimate. Even so, the RR of complications had limited impact on the cost-utility analysis results.

More importantly, the cost-effectiveness of TRT did vary according to the instrument and method used to estimate utilities. The SF-6D is a generic preference-based measure of health-related QoL based on the SF-36. The analysis of the SF-6D IPD resulted in a fairly small utility difference between TRT and SoC. When differences in SF-6D utility and complications were considered, TRT was unlikely to be cost-effective even when this utility difference was extrapolated for 10 years or for lifetime. The analysis of the BDI score mapped to the EQ-5D score resulted in a utility difference in favour of TRT that was 6.5 times larger than the difference obtained from the analysis of the SF-6D. Therefore, the cost-effectiveness of TRT improved dramatically for those scenarios where BDI-based utilities were utilised.

The SF-6D algorithm use responses to 11 of the 36 items included in the SF-36 to generate utility scores. This subset of questions covers a wide range of dimensions such as physical functioning (e.g. limitation for vigorous or moderate activities, bathing and dressing), physical role (e.g. limited in the kind of work performed), emotional role (e.g. accomplish less than you would like), pain (e.g. bodily pain, pain

interfering with usual work), mental health (e.g. felt nervous, have a lot of energy, felt downhearted or depressed) and how physical and emotional problems might have interfered with the respondent's social activities. Such items are in line with those identified as important to individuals with hypogonadism in the qualitative synthesis of the TESTES project (see [Chapter 3](#)). In such synthesis, it is highlighted that at different decision points across the timeline of living with low testosterone, several related subthemes highlight the complexity of how symptoms influence many aspects of men's lives and their treatment experiences. Such symptoms seem to affect men's daily lives in different ways, and the perceived effects of testosterone therapy were mainly discussed in the light of experienced symptoms. Responses to these items should therefore in theory reflect the expected changes experienced by individuals with hypogonadism before and after TRT.

On the other hand, the BDI evaluates key symptoms of depression such as mood, pessimism, sense of failure, self-dissatisfaction, guilt, self-dislike, self-accusation, social withdrawal and loss of libido.¹²⁵ The changes in QoL from TRT might act through these particular dimensions in individuals with hypogonadism. However, a limitation of the mapping approach provided by Grochtdreis *et al.* 2016 is its restricted generalisability. Grochtdreis *et al.* 2016 included only patients from a psychotherapeutic outpatient clinic for their mapping exercise and the application of the mapping algorithm to other patient groups is questionable.¹²⁶ Furthermore, the authors' model predictive performance in the validation samples was better for individuals with good health than for individuals with bad health. This is an indication of a systematic bias in the estimation of the mapped EQ-5D utility scores with unknown implications for the cost-effectiveness of TRT.¹⁵³ This systematic bias is a source of uncertainty that could not be contemplated through sensitivity analysis. Due to these limitations, the results using the BDI-based utilities should not be considered conclusive.

We have assumed no discontinuation of treatment for those individuals receiving TRT. The implication of this structural assumption is that all the individuals under the TRT strategy will accrue in the long term the cost of TRT, but not the benefits [in those scenarios where the relative effects (e.g. mortality and/or utility differences) were defined to last for 1 year]. Clinical expert opinion within the project team was that a small proportion of hypogonadal men might discontinue TRT during the first year of treatment, with most of them resuming treatment after a short period of time.¹⁵⁴ While the assumption of no discontinuation constitutes a limitation of our model, allowing for the cost of TRT and limited QoL benefits seems to be supported by the results of the long-term QoL IPD analysis where no differences in utilities were observed.

We have selected a cohort Markov model structure to assess the relative efficiency of TRT compared with SoC. An individual sampling model or patient-level simulation would a priori be the natural model selection when individual data and large sample size are available and when it is crucial to recreate individual care pathways to reflect heterogeneity. There were important limitations in the obtained data such as limited number of studies collecting key outcome data, similar variables being collected at very different time points or very small numbers of events happening for the trial follow-up. This persuaded us to develop a less data-demanding model that could answer the question posed and reflect the key sources of decision uncertainty.

Only one study met our inclusion criteria for the systematic literature review or economic evaluation. Arver *et al.* 2014 concluded that lifelong TRT with TU depot injection was cost-effective for men diagnosed with hypogonadism in Sweden.¹¹² There are several differences between the model reported in Arver *et al.* 2014 and our model and as such the comparison of the results for these two analyses is not straightforward. Arver *et al.* 2014 explicitly modelled fractures, major depression and T2DM as well as CV and CBV complications for those individuals developing diabetes.¹¹² The authors applied TRT-independent treatment effects favouring TRT to fractures, depression and T2DM as well as an increased risk of CV and CBV due to T2DM. Besides, significant utility reductions and increased mortality were associated with all the modelled events. We explicitly chose to follow a different approach to make the most of the analysis of the IPD obtained from the clinical investigators of the studies included

in the systematic review of clinical effectiveness. Our model incorporated estimated TRT treatment effects based on the observed differences in mortality, CV and CBV complications and HR QoL (using preference-based measures of utility) in the pooled IPD. We incorporated a direct estimate of the relative effect of TRT on CV and CBV versus SoC, rather than modelling an effect indirectly through an effect on T2DM. Finally, while we did not explicitly model fractures, depression or diabetes, their effect on QoL should be reflected, at least partially, by the general measure of utility used in our model and our results are comparable to those reported by Arver *et al.* 2014 when the BDI-based utilities were used.¹¹²

Conclusions

Our results suggest that the cost-effectiveness of TRT is dependent on its effects on all-cause mortality, and on the approach used to estimate the health state utility increment associated with TRT, which may be driven by improvements in symptoms associated with low testosterone such as sexual dysfunction and low mood. Our analysis was based on IPD encompassing the majority of existing RCTs assessing the effects of TRT in men with hypogonadism. The IPD analysis identified non-significant reductions in mortality risk during TRT when compared with placebo, though there were too few events for a reliable evaluation; inclusion and extrapolation of any putative beneficial effects of TRT on mortality are pivotal to the cost-effectiveness of TRT. We also identified that the choice of the instrument and approach to estimate QoL weights (BDI or SF-6D) was crucial to the cost-effectiveness of TRT during modelling. Furthermore, usable data to estimate QoL weights for the economic analysis were available for only a small number of studies within the collated IPD.

In summary, the ICER was below the accepted threshold for the UK (£20,000 per QALY gained) when the BDI-based utility difference lasting 10 years was used, regardless of the starting age of the population cohort (whether 40, 60 or 75 years old). However, the ICER increased above the threshold when the SF-6D utility difference was used and no difference in mortality due to TRT was assumed, regardless of the starting age of the population cohort. It is unclear whether either of these instruments fully reflects the HRQoL changes that result from TRT in hypogonadal men.

Further clarity on the CV safety of TRT in men with hypogonadism and more in-depth mapping of clinical outcomes to direct utility-based measures of HR QoL will be important to inform more robust estimates of cost-effectiveness of TRT for men with hypogonadism.

Chapter 6 Conclusions

Implications for health care

The testicular steroid hormone testosterone is critical for male sexual behaviour and physical development, so that individuals affected with MH experience significant adverse impacts on their physical and mental health. MH is an increasingly common condition due generally to increased life-expectancy, a rising prevalence of obesity and diabetes and an increasing number of survivors of adult and childhood cancers whose treatments impact on the gonadal axis. However, it is also clear that UK-wide and global prescriptions of TRT have increased at a rate exceeding any objective increments in MH disease prevalence. In a recent analysis of global prescribing data, TRT sales increased 12-fold globally from \$150 million in 2000 to \$1.8 billion in 2011. Conversely, concerns about the CV safety of TRT and over-prescribing by some clinicians have since led to a halving of TRT prescriptions in the USA since 2013.^{12,76,155} Unless gonadotropin levels are unequivocally raised or there is a clear congenital or syndromic phenotype, the diagnosis of MH is not straightforward as it requires clinicians to distinguish between infirmity resulting from testosterone deficiency (MH) and low testosterone levels resulting from infirmity (non-gonadal illness). Moreover, unlike most other hormone deficiencies, MH is not defined by low levels of serum testosterone below 95% of the population reference range. Various criteria for MH diagnosis exist – all based on a combination of clinical and biochemical features that, taken together, predict illness of a nature that is likely to be ameliorated by TRT.^{1,3,72,156} However, there remains a lack of consensus in key areas, specifically: the CV safety of TRT; which patients most benefit from TRT; the patient experience of MH and TRT; the ability of available tools to reliably measure the patient experience; the economic impact of MH and its therapy with TRT. The NIHR TestES Consortium, co-ordinated by the University of Aberdeen and Imperial College in London, is a global collaboration of principal investigators of trials conducted in nine different countries. Importantly, TestES also worked with patients to ensure that the objectives and implementation of the evidence aligned with the needs of men with MH. This project has allowed us to compile extensive IPD from double-blinded, placebo-controlled TRT monotherapy trials in men with MH, and to address questions that have been hitherto under-represented in the literature in relation to clinical and cost-effectiveness. In addition, we have collated studies outside the IPD data set to analyse the experience and perspective of men with hypogonadism.

We can make several observations and conclusions from the research presented in this assessment, which will be relevant for both patients and clinicians. First, we found no evidence that TRT significantly increases the risks of adverse CV events in men with MH. This information will be of some reassurance for clinicians and patients alike, and may be relevant for regulatory information about TRT, pending the eventual outcome of a US Food & Drugs Agency (FDA)-commissioned RCT of 6000 men powered to assess CV event risk.⁷⁶ For mortality the paucity of deaths means this question is not satisfactorily answered and we would have to wait for these events to accumulate – probably for follow-up periods of 2–5 years.

Second, our a priori analysis did not allow us to identify any patient characteristics consistently associated with the extent by which symptoms improve during TRT. For this reason, post hoc analyses were conducted; these revealed that older men and men with obesity may experience lesser symptom control during testosterone treatment compared with other men due to more severe baseline symptoms. Contrary to prior assumption and observations,²⁵ we reported that symptomatic improvements in men were experienced during TRT, regardless of baseline serum testosterone (provided it was < 12 nmol/l). Some guidelines recommend treating men with serum testosterone levels < 8 nmol/l, but others recommend higher thresholds such as < 10.4 and < 12 nmol/l.^{1,3,72,156} In addition, the recent T4DM RCT has recently reported its findings in 1007 men aged 50–74 years, with impaired glucose tolerance, waist circumference of > 95 cm, serum testosterone concentration < 14 nmol/l but without pathological

hypogonadism, given combined TRT and lifestyle intervention.¹⁵⁷ This study was excluded from TestES, which analysed the effects of TRT monotherapy, but suggests that TRT may increase diabetes remission and increase bone mineralisation in men without MH. Results of T4DM may have an unintended consequence to further wide variation in the threshold required to initiate TRT. In addition, there is no direct comparison between the effectiveness of transdermal versus injectable routes of TRT to alleviate symptoms of hypogonadism. It is, therefore, important that further evidence is generated to establish the patient characteristics and route of TRT most likely to achieve a therapeutic response in MH.

The third observation of this study is that non-sexual patient-important outcomes such as mood, cognition and social function have little corroborative RCT evidence to support their effective treatment by TRT. We also highlight the dearth of research examining the patient experience of MH outside North America and non-white populations, which may have impacted on the development of PROMs. Historically, research funders and researchers alike have prioritised investigation of the biological actions and pharmacological actions of TRT. However, we conclude that this key facet of MH of the importance to patients remains under-explored, which limits our ability to effectively identify those who would most benefit from TRT. Finally, the small number of CV events and deaths recorded during all RCTs of TRT, and lack of robust evidence on preference-based health HRQoL weights associated with the clinical benefits of TRT, imposes considerable uncertainty on any economic model comparing care pathways for men with MH.

Recommendations for future research

Based on the findings of this assessment, we make the following recommendations for future research:

- A well-designed RCT of men with hypogonadism at high risk of CV disease to inform about the safety of long-term use of TRT including its impact on mortality; an ongoing FDA-commissioned RCT⁷⁶ is likely to provide long-term safety data of TRT use, but these results may not be fully applicable to the UK or non-white ethnic groups.
- An RCT to compare symptomatic improvements from TRT versus placebo in men with hypogonadism with serum testosterone above 12 nmol/l, but below mid-point of the reference range for serum testosterone, that is, 12–15 nmol/l. A trend for TRT prescribing in these men is currently not supported by evidence, but it is becoming increasingly popular through online and private men's health clinics. Successful completion of this RCT would introduce consistency for affected patients and modify the actual definition of MH adopted by specialist societies. Ideally, generic preference-based measures of HRQoL should be collected in such a study to inform future economic evaluation studies and assess the economic implications of TRT usage for the NHS.
- An RCT comparing the effects of transdermal versus intramuscular TRT on overall satisfaction with treatment, sexual function, cognition, mood and QoL in men with hypogonadism. This would help identify characteristics such as age and diabetes diagnosis associated with preference for a particular route of TRT administration and refine current clinical practice.
- A mapping study to associate QoL instruments commonly used in existing RCT studies of MH (e.g. IIEF-15) with generic preference-based HRQoL instruments (e.g. EQ-5D) would allow researchers to incorporate existing evidence into future economic evaluations and models.
- A qualitative study exploring the experience of MH in a multiethnic patient group (across geographic locations) and co-development of a holistic symptom score for TRT response in men with hypogonadism.
- Research on route of administration of TRT that may impact the treatment effect as well as adherence (i.e. related to individual preference).
- Investigating the impact and treatment of hypogonadism in ethnically diverse patient groups.

Patient and public involvement

Development of the research question and outcome measures informed by patients' priorities, experience and preferences

The research question was formulated by the NIHR HTA research panel, which included patient members. The TestES study team including two patient members, refined the research question and agreed upon outcome measures.

Involvement of patients in the design of this study

Patient members of the TestES study team were involved as research partners in all aspects of the study including refinement of the research question, identifying areas most in need for investigation and for providing 'deciding votes and opinions' where the study team was unable to reach consensus.

Were patients involved in the conduct of the study?

A panel of patients was convened to discuss findings of the evidence syntheses, and to guide the interpretation of findings by the research team. These opinions directly fed into the interpretation of all aspects of the study. Most notably, the patient panel members helped highlight a clear divergence in the experience of men with hypogonadism of non-sexual symptoms (e.g. poor cognition and low mood) supported by qualitative data, and the paucity of supporting evidence from several large RCTs. Furthermore, patients clearly report inconsistency and uncertainty about the threshold for treating hypogonadism with TRT. These opinions have directly influenced the conclusions and recommendations of this report.

Dissemination of findings to patients and public

The TestES investigators had originally planned to hold a public conference co-led by patients and clinicians to discuss the study findings. However, the changes imposed by the COVID-19 pandemic and feedback from our patient study team members has led us to develop a website providing textual and animated video-based information resources to highlight key findings from this NIHR-funded research. The study investigators have close links with UK-based professional and patient networks (e.g. Society for Endocrinology, You and Your Hormones), to ensure that the findings from TestES have impact far beyond the duration of this evidence synthesis.

Additional information

Contributions of authors

Moira Cruickshank (<https://orcid.org/0000-0002-5182-884X>) (Research Fellow, University of Aberdeen) reviewed and summarised the current evidence on the clinical effectiveness of testosterone replacement therapy in men with low testosterone, and also assisted in screening articles for the synthesis of qualitative evidence.

Jemma Hudson (<https://orcid.org/0000-0002-6440-6419>) (Medical Statistician, University of Aberdeen) conducted the IPD and aggregate data analyses.

Rodolfo Hernández (<https://orcid.org/0000-0003-2619-8230>) (Research Fellow, University of Aberdeen) reviewed the evidence on the cost-effectiveness of testosterone replacement therapy in men with low testosterone, developed the economic model and conducted the cost-effectiveness analyses.

Magaly Aceves-Martins (<https://orcid.org/0000-0002-9441-142X>) (Research Fellow, University of Aberdeen) reviewed the qualitative evidence on the experience of men with low testosterone and conducted the PROMs analyses, and contributed to the screening and risk of bias assessment of the studies included in the quantitative synthesis.

Richard Quinton (<https://orcid.org/0000-0002-4842-8095>) (Consultant Endocrinologist and Senior Lecturer, Newcastle upon Tyne University & Hospitals NHS Foundation Trust) provided expert advice and guidance on the clinical aspects of this assessment.

Katie Gillies (<https://orcid.org/0000-0001-7890-2854>) (Reader on Research, University of Aberdeen) supervised Magaly Aceves-Martins in reviewing the qualitative evidence on the experience of men with low testosterone and conducting the PROMs analyses.

Lorna S Aucott (<https://orcid.org/0000-0001-6277-7972>) (Senior Medical Statistician, University of Aberdeen) supported Jemma Hudson in conducting the IPD and aggregate data analyses.

Charlotte Kennedy (<https://orcid.org/0000-0002-1974-6318>) (Research Assistant, University of Aberdeen) assisted Rodolfo Hernández in reviewing the evidence on the cost-effectiveness of testosterone replacement therapy in men with low testosterone, developing the economic model and conducting the cost-effectiveness analyses.

Paul Manson (<https://orcid.org/0000-0002-1405-1795>) (Information Officer, University of Aberdeen) ran literature searches and provided information support throughout the assessment.

Nicholas Oliver (<https://orcid.org/0000-0003-3525-3633>) (Consultant Diabetologist, Imperial College Healthcare NHS Trust, London) provided expert advice and guidance on the clinical aspects of this assessment.

Frederick Wu (<https://orcid.org/0000-0002-4580-8199>) (University of Manchester) provided expert advice and guidance on the clinical aspects of this assessment.

Siladitya Bhattacharya (<https://orcid.org/0000-0002-4588-356X>) (Head of School of Medicine, Medical Sciences and Nutrition, University of Aberdeen) provided expert advice and guidance on the clinical aspects of this assessment.

Waljit S Dhillon (<https://orcid.org/0000-0001-5950-4316>) (Professor of Endocrinology & Metabolism, Imperial College, London) provided expert advice and guidance on the clinical aspects of this assessment.

Channa N Jayasena (<https://orcid.org/0000-0002-2578-8223>) (Reader & NIHR Post-Doctoral Fellow, Imperial College, London) oversaw and co-ordinated all aspects of the assessment.

Miriam Brazzelli (<https://orcid.org/0000-0002-7576-6751>) (Professor of Health Services Research, University of Aberdeen) oversaw and co-ordinated all aspects of the assessment.

All authors contributed to the writing and approved the final version of this report.

Acknowledgements

The authors are grateful to Alison Avenell (Professor in Health Services research, Health Services Research Unit, University of Aberdeen, UK) and Neil Scott (Research Fellow, Medical Statistics Team, University of Aberdeen, UK), members of the Advisory Group and Graham Scotland (Senior Research Fellow, Health Services Research Unit and Health Economics Research Unit, University of Aberdeen, UK) for providing comments during preparation of the final report and to Cynthia Fraser for assisting with the development of the original literature searches.

We would like to extend our sincere thanks to our collaborators, for generously sharing their data and providing additional trial information:

Marianne Andersen (Clinical Professor, University of Southern Denmark), Shalender Bhasin (Professor, Harvard Medical School and Brigham and Women's Hospital, USA), Marielle Emmelot-Vonk (Professor, UMC Utrecht, the Netherlands), Erik Giltay (MD, PhD, University of Leiden, the Netherlands), Mathis Grossman (Professor, University of Melbourne, Australia), Kristina Groti (MD, PhD, University of Ljubljana, Slovenia), Geoff Hackett (Professor, Good Hope Hospital, Sutton Coldfield, UK), Kerry Hildreth (MD, University of Colorado Medical School), Leonard Marks (Professor, UCLA, USA), Stephen Roberts (Senior Lecturer, University of Manchester Centre for Biostatistics, UK), Richard Ross (Professor, University of Sheffield, UK), Dustin Ruff (PhD, Eli Lilly and Company), Peter Snyder (Professor, University of Pennsylvania, USA), Johan Svartberg (Professor, University Hospital of North Norway), Hui Meng Tan (Professor, Subang Jaya Medical Centre, Malaysia), Lisa Tenover (Professor, Stanford University School of Medicine, USA). Our gratitude goes also to Dina Appleby (University of Pennsylvania), Emily Gianatti (Fiona Stanley Hospital, Western Australia), Line Velling Magnussen (University of Southern Denmark) and Tom Trivison (Harvard Medical School) for taking the time to respond to our further queries.

The Health Services Research Unit and the Health Economics Research Unit at the University of Aberdeen are funded by the Chief Scientist Office of the Scottish Government Health and Social Care Directorate.

Amendments to the research protocol

The requirement for participants to be symptomatic was removed from the original version of the protocol.

Patient data statement

This work uses data provided by patients and collected by the NHS as part of their care and support. Using patient data is vital to improve health and care for everyone. There is huge potential to make

better use of information from people's patient records, to understand more about disease, develop new treatments, monitor safety and plan NHS services. Patient data should be kept safe and secure, to protect everyone's privacy, and it is important that there are safeguards to make sure that they are stored and used responsibly. Everyone should be able to find out about how patient data are used. #datasaveslives You can find out more about the background to this citation here: <https://understandingpatientdata.org.uk/data-citation>.

Data-sharing statement

All data requests should be submitted to the corresponding authors for consideration. Access to available anonymised data will be considered following review and appropriate agreement being in place.

Ethics statement

This report focuses on secondary research evidence. No primary research data were collected as part of this project and ethics approval was not required.

Information governance statement

No personal data (information that can identify a living individual) were accessed or processed during this project.

Disclosure of interests

Full disclosure of interests: Completed ICMJE forms for all authors, including all related interests, are available in the toolkit on the NIHR Journals Library report publication page at <https://doi.org/10.3310/JRYT3981>.

Primary conflicts of interest: Richard Quinton received payments or honoraria for lectures and support for attending meetings from Bayer UK. Siladitya Bhattacharya declares royalties paid to self from Cambridge University Press, speaker's honorarium and fees from Obstetrical & Gynaecological Society of Singapore paid to the University of Aberdeen and Honorarium as Editor-in-Chief, Human Reproduction Open by Oxford University Press. Waljit S Dhillon received grants or contracts from Imperial College Health Partners (IChP) and consulting fees from Imperial Consultants Ltd. Katie Gillies is a Member of the HTA Clinical Evaluation and Trials Committee 2020–2024. Lorna S Aucott was a Member of the PHR Research Funding Board 2017–2023. Nicholas Oliver is a Member of the HTA Prioritisation Committee B (in hospital) 2021–2025. Channa N Jayasena received investigator-led grant from Logixx Pharma Ltd (2016 onwards). Miriam Brazzelli is a member of the HTA Commissioning Committee 2023–2028.

Additional funding: Waljit S Dhillon (NIHR Senior Investigator Award); Channa N Jayasena (NIHR Post-Doctoral Fellowship).

Publication

Hernández R, de Silva NL, Hudson J, Cruickshank M, Quinton R, Manson P, *et al*. Cost-effectiveness of testosterone treatment utilising individual patient data from randomised controlled trials in men with low testosterone levels. *Andrology* 2024;**12**:477–486. <https://doi.org/10.1111/andr.13597>

References

1. Bhasin S, Brito JP, Cunningham GR, Hayes FJ, Hodis HN, Matsumoto AM, *et al.* Testosterone therapy in men with hypogonadism: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab* 2018;**103**(5):1715–44.
2. Handelsman DJ. Androgen physiology, pharmacology, use and misuse. In Feingold KR, Anawalt B, Boyce A, Chrousos G, de Herder WW, Dhatariya K, *et al.*, editors. *Endotext*. South Dartmouth, MA: MDText.com, Inc.; 2000.
3. Corona G, Goulis DG, Huhtaniemi I, Zitzmann M, Toppari J, Forti G, *et al.* European Academy of Andrology (EAA) guidelines on investigation, treatment and monitoring of functional hypogonadism in males: endorsing organization: European Society of Endocrinology. *Andrology* 2020;**8**(5):970–87.
4. Wu FC, Tajar A, Beynon JM, Pye SR, Silman AJ, Finn JD, *et al.*; EMAS Group. Identification of late-onset hypogonadism in middle-aged and elderly men. *N Engl J Med* 2010;**363**(2):123–35.
5. Feldman HA, Goldstein I, Hatzichristou DG, Krane RJ, McKinlay JB. Impotence and its medical and psychosocial correlates: results of the Massachusetts Male Aging Study. *J Urol* 1994;**151**(1):54–61.
6. Bacon CG, Mittleman MA, Kawachi I, Giovannucci E, Glasser DB, Rimm EB. Sexual function in men older than 50 years of age: results from the health professionals follow-up study. *Ann Intern Med* 2003;**139**(3):161–8.
7. Al-Sharefi A, Quinton R. Current national and international guidelines for the management of male hypogonadism: helping clinicians to navigate variation in diagnostic criteria and treatment recommendations. *Endocrinol Metab (Seoul)* 2020;**35**(3):526–40.
8. Kwong JCC, Krakowsky Y, Grober E. Testosterone deficiency: a review and comparison of current guidelines. *J Sex Med* 2019;**16**(6):812–20.
9. Rosen RC, Wu F, Behre HM, Porst H, Meuleman EJH, Maggi M, *et al.*; RHYME Investigators. Quality of life and sexual function benefits of long-term testosterone treatment: longitudinal results from the Registry of Hypogonadism in Men (RHYME). *J Sex Med* 2017;**14**(9):1104–15.
10. Travison TG, Vesper HW, Orwoll E, Wu F, Kaufman JM, Wang Y, *et al.* Harmonized reference ranges for circulating testosterone levels in men of four cohort studies in the United States and Europe. *J Clin Endocrinol Metab* 2017;**102**(4):1161–73.
11. Snyder PJ, Bhasin S, Cunningham GR, Matsumoto AM, Stephens-Shields AJ, Cauley JA, *et al.*; Testosterone Trials Investigators. Effects of testosterone treatment in older men. *N Engl J Med* 2016;**374**(7):611–24.
12. Basaria S, Coviello AD, Travison TG, Storer TW, Farwell WR, Jette AM, *et al.* Adverse events associated with testosterone administration. *N Engl J Med* 2010;**363**(2):109–22.
13. Gan EH, Pattman S, S HSP, Quinton R. A UK epidemic of testosterone prescribing, 2001–2010. *Clin Endocrinol (Oxf)* 2013;**79**(4):564–70.
14. Sansone A, Sansone M, Lenzi A, Romanelli F. Testosterone replacement therapy: the Emperor's new clothes. *Rejuvenation Res* 2017;**20**(1):9–14.
15. NHS England. *The 'Male Menopause' 2019*. URL: www.nhs.uk/conditions/male-menopause/ (accessed 7 May 2021).
16. Food and Drug Administration. *FDA Drug Safety Communication: FDA Cautions About Using Testosterone Products for Low Testosterone Due to Aging; Requires Labeling Change to Inform*

- of Possible Increased Risk of Heart Attack and Stroke With Use; 2015. URL: www.fda.gov/media/91048/download (accessed 7 May 2021).
17. Fernández-Balsells MM, Murad MH, Lane M, Lampropulos JF, Albuquerque F, Mullan RJ, *et al.* Clinical review 1: adverse effects of testosterone therapy in adult men: a systematic review and meta-analysis. *J Clin Endocrinol Metab* 2010;**95**(6):2560–75.
 18. Xu L, Freeman G, Cowling BJ, Schooling CM. Testosterone therapy and cardiovascular events among men: a systematic review and meta-analysis of placebo-controlled randomized trials. *BMC Med* 2013;**11**:108.
 19. Corona G, Maseroli E, Rastrelli G, Isidori AM, Sforza A, Mannucci E, Maggi M. Cardiovascular risk associated with testosterone-boosting medications: a systematic review and meta-analysis. *Expert Opin Drug Saf* 2014;**13**(10):1327–51.
 20. Borst SE, Shuster JJ, Zou B, Ye F, Jia H, Wokhlu A, Yarrow JF. Cardiovascular risks and elevation of serum DHT vary by route of testosterone administration: a systematic review and meta-analysis. *BMC Med* 2014;**12**:211.
 21. Seftel AD, Kathrins M, Niederberger C. Critical update of the 2010 Endocrine Society clinical practice guidelines for male hypogonadism: a systematic analysis. *Mayo Clin Proc* 2015;**90**(8):1104–15.
 22. Albert SG, Morley JE. Testosterone therapy, association with age, initiation and mode of therapy with cardiovascular events: a systematic review. *Clin Endocrinol (Oxf)* 2016;**85**(3):436–43.
 23. Guo C, Gu W, Liu M, Peng BO, Yao X, Yang B, Zheng J. Efficacy and safety of testosterone replacement therapy in men with hypogonadism: a meta-analysis study of placebo-controlled trials. *Exp Ther Med* 2016;**11**(3):853–63.
 24. Huo S, Scialli AR, McGarvey S, Hill E, Tügetimur B, Hogenmiller A, *et al.* Treatment of men for 'low testosterone': a systematic review. *PLOS ONE* 2016;**11**(9):e0162480.
 25. Corona G, Rastrelli G, Morgentaler A, Sforza A, Mannucci E, Maggi M. Meta-analysis of results of testosterone therapy on sexual function based on international index of erectile function scores. *Eur Urol* 2017;**72**(6):1000–11.
 26. Alexander GC, Iyer G, Lucas E, Lin D, Singh S. Cardiovascular risks of exogenous testosterone use among men: a systematic review and meta-analysis. *Am J Med* 2017;**130**(3):293–305.
 27. Szeinbach SL, Seoane-Vazquez E, Summers KH. Development of a men's Preference for Testosterone Replacement Therapy (P-TRT) instrument. *Patient Prefer Adherence* 2012;**6**:631–41.
 28. Mascarenhas A, Khan S, Sayal R, Knowles S, Gomes T, Moore JE. Factors that may be influencing the rise in prescription testosterone replacement therapy in adult men: a qualitative study. *Aging Male* 2016;**19**(2):90–5.
 29. Dunning TL, Ward GM. Testosterone replacement therapy – perceptions of recipients and partners. *J Adv Nurs* 2004;**47**(5):467–74.
 30. Hackett G, Kirby M, Edwards D, Jones TH, Wylie K, Ossei-Gerning N, *et al.* British Society for Sexual Medicine guidelines on adult testosterone deficiency, with statements for UK practice. *J Sex Med* 2017;**14**(12):1504–23.
 31. Bhasin S, Cunningham GR, Hayes FJ, Matsumoto AM, Snyder PJ, Swerdloff RS, Montori VM; Task Force, Endocrine Society. Testosterone therapy in men with androgen deficiency syndromes: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab* 2010;**95**(6):2536–59.

32. Hicks KA, Mahaffey KW, Mehran R, Nissen SE, Wiviott SD, Dunn B, *et al.*; Standardized Data Collection for Cardiovascular Trials Initiative (SCTI). 2017 Cardiovascular and stroke endpoint definitions for clinical trials. *J Am Coll Cardiol* 2018;**71**(9):1021–34.
33. StataCorp. *Stata Statistical Software: Release 16*. College Station, TX: StataCorp LLC; 2019.
34. Higgins JP, Green S. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0*; 2011. URL: <http://handbook-5-1.cochrane.org/> (accessed 29 April 2021).
35. Maruish M. *User's manual for the SF-36v2 Health Survey*. 3rd edn. Lincoln, RI: QualityMetric Incorporated; 2011.
36. Efthimiou O. Practical guide to the meta-analysis of rare events. *Evid Based Ment Health* 2018;**21**(2):72–6.
37. Riley RD, Debray TPA, Fisher D, Hattle M, Marlin N, Hoogland J, *et al.* Individual participant data meta-analysis to examine interactions between treatment effect and participant-level covariates: statistical recommendations for conduct and planning. *Stat Med* 2020;**39**(15):2115–37.
38. Amory JK, Watts NB, Easley KA, Sutton PR, Anawalt BD, Matsumoto AM, *et al.* Exogenous testosterone or testosterone with finasteride increases bone mineral density in older men with low serum testosterone. *J Clin Endocrinol Metab* 2004;**89**(2):503–10.
39. Basaria S, Harman SM, Travison TG, Hodis H, Tsitouras P, Budoff M, *et al.* Effects of testosterone administration for 3 years on subclinical atherosclerosis progression in older men with low or low-normal testosterone levels: a randomized clinical trial. *JAMA* 2015;**314**(6):570–81.
40. Hildreth KL, Barry DW, Moreau KL, Vande Griend J, Meacham RB, Nakamura T, *et al.* Effects of testosterone and progressive resistance exercise in healthy, highly functioning older men with low-normal testosterone levels. *J Clin Endocrinol Metab* 2013;**98**(5):1891–900.
41. Marks LS, Mazer NA, Mostaghel E, Hess DL, Dorey FJ, Epstein JI, *et al.* Effect of testosterone replacement therapy on prostate tissue in men with late-onset hypogonadism – a randomized controlled trial. *JAMA* 2006;**296**(19):2351–61.
42. 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):1612–27.
43. Merza Z, Blumsohn A, Mah PM, Meads DM, McKenna SP, Wylie K, *et al.* Double-blind placebo-controlled study of testosterone patch therapy on bone turnover in men with borderline hypogonadism. *Int J Androl* 2006;**29**(3):381–91.
44. Srinivas-Shankar U, Roberts SA, Connolly MJ, O'Connell MD, Adams JE, Oldham JA, Wu FCW. Effects of testosterone on muscle strength, physical function, body composition, and quality of life in intermediate-frail and frail elderly men: a randomized, double-blind, placebo-controlled study. *J Clin Endocrinol Metab* 2010;**95**(2):639–50.
45. Emmelot-Vonk MH, Verhaar HJJ, Pour HRN, Aleman A, Lock T, Bosch J, *et al.* Effect of testosterone supplementation on functional mobility, cognition, and other parameters in older men – a randomized controlled trial. *JAMA* 2008;**299**(1):39–52.
46. Gianatti EJ, Dupuis P, Hoermann R, Zajac JD, Grossmann M. Effect of testosterone treatment on constitutional and sexual symptoms in men with type 2 diabetes in a randomized, placebo-controlled clinical trial. *J Clin Endocrinol Metab* 2014;**99**(10):3821–8.
47. Giltay EJ, Tishova YA, Mskhalaya GJ, Gooren LJ, 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):2572–82.

48. Groti K, Žuran I, Antonič B, Foršnarič L, Pfeifer M. The impact of testosterone replacement therapy on glycemic control, vascular function, and components of the metabolic syndrome in obese hypogonadal men with type 2 diabetes. *Aging Male* 2018;**21**(3):158–69.
49. Ho CC, Tong SF, Low WY, Ng CJ, Khoo EM, Lee VK, *et al.* A randomized, double-blind, placebo-controlled trial on the effect of long-acting testosterone treatment as assessed by the Aging Male Symptoms scale. *BJU Int* 2012;**110**(2):260–5.
50. Magnussen LV, Glinborg D, Hermann P, Hougaard DM, Højlund K, Andersen M. Effect of testosterone on insulin sensitivity, oxidative metabolism and body composition in aging men with type 2 diabetes on metformin monotherapy. *Diabetes Obes Metab* 2016;**18**(10):980–9.
51. 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):378–87.
52. Brock G, Heiselman D, Maggi M, Kim SW, Rodríguez Vallejo JM, Behre HM, *et al.* Effect of testosterone solution 2% on testosterone concentration, sex drive and energy in hypogonadal men: results of a placebo controlled study. *J Urol* 2016;**195**(3):699–705.
53. Borst SE, Yarrow JF, Conover CF, Nseyo U, Meuleman JR, Lipinska JA, *et al.* Musculoskeletal and prostate effects of combined testosterone and finasteride administration in older hypogonadal men: a randomized, controlled trial. *Am J Physiol Endocrinol Metab* 2014;**306**(4):E433–42.
54. 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):421–30.
55. Dhindsa S, Ghanim H, Batra M, Kuhadiya ND, Abuaysheh S, Sandhu S, *et al.* Insulin resistance and inflammation in hypogonadotropic hypogonadism and their reduction after testosterone replacement in men with type 2 diabetes. *Diabetes Care* 2016;**39**(1):82–91.
56. Dias JP, Melvin D, Simonsick EM, Carlson O, Shardell MD, Ferrucci L, *et al.* Effects of aromatase inhibition vs. testosterone in older men with low testosterone: randomized-controlled trial. *Andrology* 2016;**4**(1):33–40.
57. Kaufman JM, Miller MG, Garwin JL, Fitzpatrick S, McWhirter C, Brennan JJ. Efficacy and safety study of 1.62% testosterone gel for the treatment of hypogonadal men. *J Sex Med* 2011;**8**(7):2079–89.
58. Kenny AM, Kleppinger A, Annis K, Rathier M, Browner B, Judge JO, McGee D. Effects of transdermal testosterone on bone and muscle in older men with low bioavailable testosterone levels, low bone mass, and physical frailty. *J Am Geriatr Soc* 2010;**58**(6):1134–43.
59. Steidle C, Schwartz S, Jacoby K, Sebree T, Smith T, Bachand RAA. 2500 testosterone gel normalizes androgen levels in aging males with improvements in body composition and sexual function. *J Clin Endocrinol Metab* 2003;**88**(6):2673–81.
60. Aversa A, Bruzziches R, Francomano D, Rosano G, Isidori AM, Lenzi A, Spera G. Effects of testosterone undecanoate on cardiovascular risk factors and atherosclerosis in middle-aged men with late-onset hypogonadism and metabolic syndrome: results from a 24-month, randomized, double-blind, placebo-controlled study. *J Sex Med* 2010a;**7**(10):3495–503.
61. Aversa A, Bruzziches R, Francomano D, Spera G, Lenzi A. Efficacy and safety of two different testosterone undecanoate formulations in hypogonadal men with metabolic syndrome. *J Endocrinol Invest* 2010b;**33**(11):776–83.
62. 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):641–6.

63. Basurto L, Zarate A, Gomez R, Vargas C, Saucedo R, Galvan R. Effect of testosterone therapy on lumbar spine and hip mineral density in elderly men. *Aging Male* 2008;**11**(3):140–5.
64. Chiang HS, Hwang TI, Hsui YS, Lin YC, Chen HE, Chen GC, Liao CH. Transdermal testosterone gel increases serum testosterone levels in hypogonadal men in Taiwan with improvements in sexual function. *Int J Impot Res* 2007;**19**(4):411–7.
65. Clague JE, Wu FC, Horan MA. Difficulties in measuring the effect of testosterone replacement therapy on muscle function in older men. *Int J Androl* 1999;**22**(4):261–5.
66. Morales A, Black A, Emerson L, Barkin J, Kuzmarov I, Day A. Androgens and sexual function: a placebo-controlled, randomized, double-blind study of testosterone vs. dehydroepiandrosterone in men with sexual dysfunction and androgen deficiency. *Aging Male* 2009;**12**(4):104–12.
67. Wang YJ, Zhan JK, Huang W, Wang Y, Liu Y, Wang S, *et al.* Effects of low-dose testosterone undecanoate treatment on bone mineral density and bone turnover markers in elderly male osteoporosis with low serum testosterone. *Int J Endocrinol* 2013;**2013**:570413.
68. Behre HM, Tammela TL, Arver S, Tolrá JR, Bonifacio V, Lamche M, *et al.*; European Testogel® Study Team. A randomized, double-blind, placebo-controlled trial of testosterone gel on body composition and health-related quality-of-life in men with hypogonadal to low-normal levels of serum testosterone and symptoms of androgen deficiency over 6 months with 12 months open-label follow-up. *Aging Male* 2012;**15**(4):198–207.
69. Jones TH, Arver S, Behre HM, Buvat J, Meuleman E, Moncada I, *et al.*; TIMES2 Investigators. Testosterone replacement in hypogonadal men with Type 2 diabetes and/or metabolic syndrome (the TIMES2 Study). *Diabetes Care*. 2011;**34**(4):828–37.
70. Paduch DA, Polzer PK, Ni X, Basaria S. Testosterone replacement in androgen-deficient men with ejaculatory dysfunction: a randomized controlled trial. *J Clin Endocrinol Metab* 2015;**100**(8):2956–62.
71. Bayer AG. *Testosterone Conversion Tool*; 2021. URL: www.nebido.com/hcp/tools/testosterone-unit-conversion (accessed 30 April 2021).
72. Diem SJ, Greer NL, MacDonald R, McKenzie LG, Dahm P, Ercan-Fang N, *et al.* Efficacy and safety of testosterone treatment in men: an evidence report for a clinical practice guideline by the American College of Physicians. *Ann Intern Med* 2020;**172**(2):105–18.
73. Bachman E, Feng R, Travison T, Li M, Olbina G, Ostland V, *et al.* Testosterone suppresses hepcidin in men: a potential mechanism for testosterone-induced erythrocytosis. *J Clin Endocrinol Metab* 2010;**95**(10):4743–7.
74. Hackett G, Cole N, Mulay A, Strange RC, Ramachandran S. Long-term testosterone therapy in type 2 diabetes is associated with reduced mortality without improvement in conventional cardiovascular risk factors. *BJU Int* 2019;**123**(3):519–29.
75. Muraleedharan V, Marsh H, Kapoor D, Channer KS, Jones TH. Testosterone deficiency is associated with increased risk of mortality and testosterone replacement improves survival in men with type 2 diabetes. *Eur J Endocrinol* 2013;**169**(6):725–33.
76. National Library of Medicine. *A Study to Evaluate the Effect of Testosterone Replacement Therapy (TRT) on the Incidence of Major Adverse Cardiovascular Events (MACE) and Efficacy Measures in Hypogonadal Men (TRAVERSE)*; 2018. URL: www.clinicaltrials.gov/ct2/show/NCT03518034?term=NCT03518034&draw=2&rank=1 (accessed August 2021).
77. Miller WR. Qualitative research findings as evidence: utility in nursing practice. *Clin Nurse Spec* 2010;**24**(4):191–3.
78. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Med Res Methodol* 2008;**8**:45.

79. Critical Appraisal Skills Programme. *CASP Qualitative Check List*. Oxford: Critical Appraisal Skills Programme; 2018. URL: https://casp-uk.b-cdn.net/wp-content/uploads/2018/03/CASP-Qualitative-Checklist-2018_fillable_form.pdf (accessed May 2021).
80. Lewin S, Glenton C, Munthe-Kaas H, Carlsen B, Colvin CJ, Gülmezoglu M, *et al*. Using qualitative evidence in decision making for health and social interventions: an approach to assess confidence in findings from evidence syntheses (GRADE-CERQual). *PLOS Med* 2015;**12**(10):e1001895.
81. Hayes RP, Henne J, Kinchen KS. Establishing the content validity of the Sexual Arousal, Interest, and Drive Scale and the hypogonadism energy diary. *Int J Clin Pract* 2015;**69**(4):454–65.
82. Gelhorn HL, Vernon MK, Stewart KD, Miller MG, Brod M, Althof SE, *et al*. Content validity of the hypogonadism impact of symptoms questionnaire (HIS-Q): a patient-reported outcome measure to evaluate symptoms of hypogonadism. *Patient* 2016;**9**(2):181–90.
83. Gelhorn HL, Bodhani AR, Wahala LS, Sexton C, Landrian A, Miller MG, *et al*. Development of the Hypogonadism Impact of Symptoms Questionnaire Short Form: qualitative research. *J Sex Med* 2016b;**13**(11):1729–36.
84. Rosen RC, Araujo AB, Connor MK, Elstad EA, McGraw SA, Guay AT, *et al*. Assessing symptoms of hypogonadism by self-administered questionnaire: qualitative findings in patients and controls. *Aging Male* 2009;**12**(2–3):77–85.
85. Chambers SK, Chung E, Wittert G, Hyde MK. Erectile dysfunction, masculinity, and psychosocial outcomes: a review of the experiences of men after prostate cancer treatment. *Transl Androl Urol* 2017;**6**(1):60–8.
86. Wentzell E. How did erectile dysfunction become ‘natural?’ A review of the critical social scientific literature on medical treatment for male sexual dysfunction. *J Sex Res* 2017;**54**(4–5):486–506.
87. King AJ, Evans M, Moore TH, Paterson C, Sharp D, Persad R, Huntley AL. Prostate cancer and supportive care: a systematic review and qualitative synthesis of men’s experiences and unmet needs. *Eur J Cancer Care (Engl)* 2015;**24**(5):618–34.
88. Zaider T, Manne S, Nelson C, Mulhall J, Kissane D. Loss of masculine identity, marital affection, and sexual bother in men with localized prostate cancer. *J Sex Med* 2012;**9**(10):2724–32.
89. Seidler ZE, Dawes AJ, Rice SM, Oliffe JL, Dhillon HM. The role of masculinity in men’s help-seeking for depression: a systematic review. *Clin Psychol Rev* 2016;**49**:106–18.
90. Handelsman DJ. Pharmacoepidemiology of testosterone prescribing in Australia, 1992–2010. *Med J Aust* 2012;**196**(10):642–5.
91. Bhugra D, Silva PD. Sexual dysfunction across cultures. *Int Rev Psychiatry* 1993;**5**(2–3):243–52.
92. Black N. Patient reported outcome measures could help transform healthcare. *BMJ* 2013;**346**:f167.
93. Langham S, Maggi M, Schulman C, Quinton R, Uhl-Hochgraeber K. Health-related quality of life instruments in studies of adult men with testosterone deficiency syndrome: a critical assessment. *J Sex Med* 2008;**5**(12):2842–52.
94. Macefield RC, Jacobs M, Korfage IJ, Nicklin J, Whistance RN, Brookes ST, *et al*. Developing core outcomes sets: methods for identifying and including patient-reported outcomes (PROs). *Trials* 2014;**15**:49.
95. Gillies K, Duthie A, Cotton S, Campbell MK. Patient reported measures of informed consent for clinical trials: a systematic review. *PLOS ONE* 2018;**13**(6):e0199775.

96. Cruickshank M, Newlands R, Blazeby J, Ahmed I, Bekheit M, Brazzelli M, *et al.* Identification and categorisation of relevant outcomes for symptomatic uncomplicated gallstone disease: in-depth analysis to inform the development of a core outcome set. *BMJ Open* 2021;**11**(6):e045568.
97. Potter WJ, Levine-Donnerstein D. Rethinking validity and reliability in content analysis. *J Appl Commun Res* 1999;**27**(3):258–84.
98. World Health Organization. *International Classification of Functioning, Disability and Health (ICF)*; 2001. URL: www.who.int/standards/classifications/international-classification-of-functioning-disability-and-health (accessed 4 August 2020).
99. Morley JE, Charlton E, Patrick P, Kaiser FE, Cadeau P, McCreedy D, Perry HM. Validation of a screening questionnaire for androgen deficiency in aging males. *Metabolism* 2000;**49**(9):1239–42.
100. Heinemann L, Saad F, Pöllänen P. Measurement of quality of life specific for aging males. In Schneider HPG, editor. *Hormone Replacement Therapy and Quality of Life*. London: Parthenon; 2002.
101. Heinemann LA, Saad F, Zimmermann T, Novak A, Myon E, Badia X, *et al.* The Aging Males' Symptoms (AMS) scale: update and compilation of international versions. *Health Qual Life Outcomes* 2003;**1**:15.
102. Heinemann LAJ, Saad F, Thiele K, Wood-Dauphinee S. The Aging Males' Symptoms rating scale: cultural and linguistic validation into English. *Aging Male* 2001;**4**(1):14–22.
103. Heinemann LAJ, Zimmermann T, Vermeulen A, Thiel C, Hummel W. A new 'aging males' symptoms' rating scale. *Aging Male* 1999;**2**(2):105–14.
104. Corona G, Mannucci E, Petrone L, Balercia G, Fisher AD, Chiarini V, *et al.* ANDROTEST: a structured interview for the screening of hypogonadism in patients with sexual dysfunction. *J Sex Med* 2006;**3**(4):706–15.
105. McMillan CV, Bradley C, Giannoulis M, Martin F, Sönksen PH. Preliminary development of a new individualised questionnaire measuring quality of life in older men with age-related hormonal decline: the A-RHDQoL. *Health Qual Life Outcomes* 2003;**1**:51.
106. Gelhorn HL, Dashiell-Aje E, Miller MG, DeRogatis LR, Dobs A, Seftel AD, *et al.* Psychometric evaluation of the hypogonadism impact of symptoms questionnaire. *J Sex Med* 2016c;**13**(11):1737–49.
107. Smith KW, Feldman HA, McKinlay JB. Construction and field validation of a self-administered screener for testosterone deficiency (hypogonadism) in ageing men. *Clin Endocrinol (Oxf)* 2000;**53**(6):703–11.
108. Morales A, Lunenfeld B; International Society for the Study of the Aging Male. Investigation, treatment and monitoring of late-onset hypogonadism in males. Official recommendations of ISSAM. International society for the study of the aging male. *Aging Male* 2002;**5**(2):74–86.
109. Bradley C. *Age-Related Hormone Deficiency-Dependent Quality of Life (A-RHDQoL)*. Egham, UK: Health Psychology Research Unit, University of London; 2001. URL: www.healthpsychologyresearch.com/find-a-questionnaire/age-related-hormone-deficiency-dependent-quality-life-rhdqol (accessed 11 August 2020).
110. Wiltink J, Beutel ME, Brähler E, Weidner W. Hypogonadism-related symptoms: development and evaluation of an empirically derived self-rating instrument (HRS 'Hypogonadism Related Symptom Scale'). *Andrologia* 2009;**41**(5):297–304.

REFERENCES

111. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, *et al.*; CHEERS Task Force. Consolidated health economic evaluation reporting standards (CHEERS) statement. *Value Health* 2013;**16**(2):e1–5.
112. Arver S, Luong B, Fraschke A, Ghatnekar O, Stanicic S, Gulyev D, Müller E. Is testosterone replacement therapy in males with hypogonadism cost-effective? An analysis in Sweden. *J Sex Med* 2014;**11**(1):262–72.
113. TreeAge Software LLC. *TreeAge Pro*. Williamstown, MA: TreeAge Software LLC; 2020.
114. National Institute for Health and Care Excellence. *Guide to the methods of technology appraisal [PMG9] 2013 updated 4 March 2020*. www.nice.org.uk/process/pmg9/chapter/foreword (accessed August 2021).
115. British Heart Foundation. *Heart and Circulatory Disease Statistics 2020*; 2020. www.bhf.org.uk/what-we-do/our-research/heart-statistics/heart-statistics-publications/cardiovascular-disease-statistics-2020 (accessed August 2021).
116. Office for National Statistics. *National Life Tables: UK*; 2020. www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/datasets/nationallifetablesunitedkingdomreferencetables (accessed August 2021).
117. National Institute for Health and Care Excellence. *Acute Coronary Syndromes Cost-effectiveness Analysis: Which Dual Antiplatelet Therapy is Most Cost Effective for Managing Unstable Angina or NSTEMI or for Managing STEMI in Adults Undergoing PCI? (NG185)*. London/Manchester: National Institute for Health and Care Excellence; 2020. www.nice.org.uk/guidance/ng185/evidence/health-economic-analysis-for-dual-antiplatelet-therapy-pdf-8903834893 (accessed 9 September 2020).
118. Smolina K, Wright FL, Rayner M, Goldacre MJ. Long-term survival and recurrence after acute myocardial infarction in England, 2004 to 2010. *Circ Cardiovasc Qual Outcomes* 2012;**5**(4):532–40.
119. Brønnum-Hansen H, Davidsen M, Thorvaldsen P; Danish MONICA Study Group. Long-term survival and causes of death after stroke. *Stroke* 2001;**32**(9):2131–6.
120. Health Economics Research Centre. *HERC Database of Mapping Studies*; 2020. www.herc.ox.ac.uk/downloads/herc-database-of-mapping-studies (accessed August 2021).
121. University of Sheffield: School of Health and Related Research. *Measuring and Valuing Health*. www.sheffield.ac.uk/scharr/research/themes/valuing-health#SF-6D (accessed 2 September 2021).
122. van den Berg B. SF-6D population norms. *Health Econ* 2012;**21**(12):1508–12.
123. Ara R, Brazier JE. Populating an economic model with health state utility values: moving toward better practice. *Value Health* 2010;**13**(5):509–18.
124. Agledahl I, Hansen JB, Svartberg J. Impact of testosterone treatment on postprandial triglyceride metabolism in elderly men with subnormal testosterone levels. *Scand J Clin Lab Invest* 2008;**68**(7):641–8.
125. Sajatovic M, Chen P, Young R. Rating Scales in bipolar disorder. In Tohen M, Bowden CL, Nierenberg AA, Geddes JR, editors. *Clinical Trial Design Challenges in Mood Disorders*. Amsterdam: Academic Press; 2015.
126. Grochtdreis T, Brettschneider C, Hajek A, Schierz K, Hoyer J, Koenig HH. Mapping the Beck depression inventory to the EQ-5D-3L in patients with depressive disorders. *J Ment Health Policy Econ* 2016;**19**(2):79–89.

127. Stolk EA, Busschbach JJ, Caffa M, Meuleman EJ, Rutten FF. Cost utility analysis of sildenafil compared with papaverine–phenolamine injections. *BMJ* 2000;**320**(7243):1165–8.
128. Heald AH, Stedman M, Whyte M, Livingston M, Albanese M, Ramachandran S, Hackett G. Lessons learnt from the variation across 6741 family/general practices in England in the use of treatments for hypogonadism. *Clin Endocrinol (Oxf)* 2021;**94**(5):827–36.
129. Electronic Medicines Compendium. *Testogel 16.2 mg/g gel*; 2015. URL: www.medicines.org.uk/emc/product/8919#gref (accessed August 2021).
130. Nottinghamshire Area Prescribing Committee. *Testosterone Replacement Therapy for Adult Male Hypogonadism*; 2020. URL: www.nottsapc.nhs.uk/media/1261/testosterone-info-sheet.pdf?UNLID= (accessed August 2021).
131. Bell C, Hadi MA, Khanal S, Paudyal V. Prescribing patterns and costs associated with erectile dysfunction drugs in England: a time trend analysis. *BJGP Open* 2021;**5**(2):bjgpopen20X101145.
132. Medicines Complete. *British National Formulary*. London: Royal Pharmaceutical Society; 2018. URL: www.medicinescomplete.com/mc/bnf/current/index.htm (accessed 4 March 2020).
133. Larsen K. *Graphreader*. URL: www.graphreader.com/ (accessed August 2021).
134. Morales A, Johnston B, Heaton JP, Lundie M. Testosterone supplementation for hypogonadal impotence: assessment of biochemical measures and therapeutic outcomes. *J Urol* 1997;**157**(3):849–54.
135. National Clinical Guideline Centre. *Appendix L: Cost-effectiveness Analysis: Low-intensity, Medium-intensity and High-intensity Statin Treatment for the Primary and Secondary Prevention of CVD (CG181)*. London: National Clinical Guideline Centre; 2014. URL: www.nice.org.uk/guidance/cg181/evidence/lipid-modification-update-appendices-pdf-243786638 (accessed 9 September 2020).
136. National Institute for Health and Care Excellence. *Peripheral arterial disease: diagnosis and management [CG147]*; 2012. URL: www.nice.org.uk/guidance/cg147 (accessed August 2021).
137. National Institute for Health and Care Excellence. *Abdominal Aortic Aneurysm: Diagnosis and Management. NICE Guideline: Health Economics (NG156)*. London/Manchester: National Institute for Health and Care Excellence; 2020. URL: www.nice.org.uk/guidance/ng156/evidence/y-health-economics-appendix-pdf-255167681407 (accessed 9 September 2020).
138. National Institute for Health and Care Excellence. *Atrial Fibrillation and Heart Valve Disease: Self-monitoring Coagulation Status Using Point-of-care Coagulometers (the CoaguChek XS system) [DG14] 2014 Updated*; 2017. URL: www.nice.org.uk/guidance/dg14 (accessed August 2021).
139. Perkins GD, Ji C, Achana F, Black JJ, Charlton K, Crawford J, et al. Adrenaline to improve survival in out-of-hospital cardiac arrest: the PARAMEDIC2 RCT. *Health Technol Assess* 2021;**25**(25):1–166.
140. NHS England. *National Schedule of NHS Costs*; 2020. URL: www.england.nhs.uk/wp-content/uploads/2021/06/National_Schedule_of_NHS_Costs_FY1920.xlsx (accessed August 2021).
141. Curtis L, Burns A. *Unit Costs of Health and Social Care 2020*; 2020. URL: www.pssru.ac.uk/project-pages/unit-costs/unit-costs-2020/ (accessed July 2021).
142. NHS England. *Arrhythmia 2018*. URL: www.nhs.uk/conditions/arrhythmia/ (accessed August 2021).
143. Burdett P, Lip GYH. Atrial fibrillation in the United Kingdom: predicting costs of an emerging epidemic recognising and forecasting the cost drivers of atrial fibrillation-related costs. *Eur Heart J Qual Care Clin Outcomes* 2020;**8**(2):187–94. <https://doi.org/10.1093/ehjqcco/qcaa093>.

144. Walker S, Asaria M, Manca A, Palmer S, Gale CP, Shah AD, *et al.* Long-term healthcare use and costs in patients with stable coronary artery disease: a population-based cohort using linked health records (CALIBER). *Eur Heart J Qual Care Clin Outcomes* 2016;**2**(2):125–40.
145. Danese MD, Gleeson M, Kutikova L, Griffiths RI, Azough A, Khunti K, *et al.* Estimating the economic burden of cardiovascular events in patients receiving lipid-modifying therapy in the UK. *BMJ Open* 2016;**6**(8):e011805.
146. NHS England. *Why It's Done - Aortic Valve Replacement*; 2018. URL: www.nhs.uk/conditions/aortic-valve-replacement/whyitsdone/ (accessed August 2021).
147. National Institute for Health and Care Excellence. *Acute Heart Failure: Diagnosis and Management [CG187]*; 2014. URL: www.nice.org.uk/guidance/cg187 (accessed August 2021).
148. NHS England. *2019/20 National Cost Collection Data*; 2020. URL: www.england.nhs.uk/national-cost-collection/#ncc1819 (accessed 26 November 2020).
149. National Institute for Health and Care Excellence. *Appendix K: Cost-Effectiveness Analysis: Supervised Exercise Compared to Unsupervised Exercise for the Treatment of People with Intermittent Claudication [NG147]*. London/Manchester: National Institute for Health and Care Excellence; 2012. URL: www.nice.org.uk/guidance/cg147/evidence/appendices-an-pdf-186865022 (accessed 25 August 2021).
150. Tappenden P, Chilcott JB. Avoiding and identifying errors and other threats to the credibility of health economic models. *PharmacoEcon* 2014;**32**(10):967–79.
151. ClinRisk Ltd. *QRISK®3-2018 Risk Calculator*; 2018. URL: <https://qrisk.org/three/index.php> (accessed August 2021).
152. Briggs A, Claxton M, Sculpher M. *Decision Modelling for Health Economic Evaluation*. Oxford: Oxford University Press; 2006.
153. Longworth L, Rowen D. *NICE DSU Technical Support Document 10: The Use of Mapping Methods to Estimate Health State Utility Values*. Sheffield: NICE Decision Support Unit, University of Sheffield; 2011. URL: <http://nicedsu.org.uk/wp-content/uploads/2016/03/TSD-10-mapping-FINAL.pdf> (accessed 24 March 2020).
154. Park H, Park J, Kim K, Kang B, Baek S, Park N. Discontinuation of testosterone replacement therapy in patients with late-onset hypogonadism: a 10-year observational study. *Eur Urol Open Sci* 2020;**19**:e131.
155. Budoff MJ, Ellenberg SS, Lewis CE, Mohler ER 3rd, Wenger NK, Bhasin S, *et al.* Testosterone treatment and coronary artery plaque volume in older men with low testosterone. *JAMA* 2017;**317**(7):708–16.
156. Yeap BB, Grossmann M, McLachlan RI, Handelsman DJ, Wittert GA, Conway AJ, *et al.* Endocrine Society of Australia position statement on male hypogonadism (part 2): treatment and therapeutic considerations. *Med J Aust* 2016;**205**(5):228–31.
157. Wittert G, Bracken K, Robledo KP, Grossmann M, Yeap BB, Handelsman DJ, *et al.* Testosterone treatment to prevent or revert type 2 diabetes in men enrolled in a lifestyle programme (T4DM): a randomised, double-blind, placebo-controlled, 2-year, phase 3b trial. *Lancet Diabetes Endocrinol* 2021;**9**(1):32–45.

Appendix 1 Ovid (MEDLINE and Embase) search strategy

Clinical review

Database: Ovid Embase <1980 to 2018 Week 35>, Ovid MEDLINE® and Epub Ahead of Print, In-process and Other Non-indexed Citations and Daily <1946 to August 24, 2018>

Date of search 27 August 2018

- 1 exp androgens/tu use ppez (7642)
- 2 hormone replacement therapy/ use ppez (9272)
- 3 2 and (men or androgen? or testosterone).af. (2597)
- 4 Androgen Therapy/ use emez (5220)
- 5 (androgen replacement therapy or art).tw,kw. (186233)
- 6 testosterone.tw,kw. (161015)
- 7 or/1,3-6 (353317)
- 8 exp Erectile Dysfunction/ use ppez (17850)
- 9 exp impotence/ use emez (38953)
- 10 Sexual Dysfunction, Physiological/ (22063)
- 11 testosterone/df (1227)
- 12 Libido/ use ppez (4538)
- 13 Libido Disorder/ use emez (5704)
- 14 Hypogonadism/ (21649)
- 15 (erectile adj3 dysfunction).tw,kw. (37580)
- 16 (libido adj3 (low\$ or decreas\$ or reduc\$ or loss)).tw,kw. (4553)
- 17 (impotence or impotent).tw,kw. (14427)
- 18 hypogonad\$.tw,kw. (28070)
- 19 (low\$ adj3 testosterone).tw. (11853)
- 20 (deficien\$ adj3 (androgen or gonad\$ or testosterone)).tw. (8153)
- 21 (insuffic\$ adj3 (androgen or gonad\$ or testosterone)).tw. (953)
- 22 (kallman or klinefetter).tw. (181)
- 23 or/8-22 (140217)
- 24 7 and 23 (30320)
- 25 exp clinical trial/ use emez (1309842)
- 26 randomized controlled trial.pt. (467661)
- 27 controlled clinical trial.pt. (92614)
- 28 randomization/ use emez (78791)
- 29 randomi?ed.ab. (1210469)
- 30 placebo.ab. (452150)
- 31 drug therapy.fs. (5417963)
- 32 randomly.ab. (676570)
- 33 trial.ab. (1049510)
- 34 groups.ab. (4245410)
- 35 or/25-34 (10765717)
- 36 exp animals/ not humans/ (15654408)
- 37 nonhuman/ not human/ (4188739)
- 38 35 not (36 or 37) (7089138)
- 39 24 and 38 (9708)

- 40 limit 39 to english language (8722)
- 41 limit 40 to (english language and yr="1992 -Current") (7714)
- 42 41 not ((women not men) or (female not male)).tw. (6951)
- 43 41 and male/ (5690)
- 44 42 or 43 (7041)

Qualitative review

Database: Embase <1980 to 2018 Week 36>, Ovid MEDLINE® and Epub Ahead of Print, In-process & Other Non-indexed Citations and Daily <1946 to September 04, 2018>

Date of Search: 5 September 2018

- 1 exp androgens/tu use ppez (7645)
- 2 hormone replacement therapy/ use ppez (9277)
- 3 2 and (men or androgen? or testosterone).af. (2599)
- 4 Androgen Therapy/ use emez (5233)
- 5 androgen replacement therapy.tw,kw. (800)
- 6 testosterone.tw,kw. (161133)
- 7 or/1,3-6 (168295)
- 8 exp Erectile Dysfunction/ use ppez (17857)
- 9 exp impotence/ use emez (38986)
- 10 Sexual Dysfunction, Physiological/ (22085)
- 11 testosterone/df (1228)
- 12 Libido/ use ppez (4540)
- 13 Libido Disorder/ use emez (5707)
- 14 Hypogonadism/ (21677)
- 15 (erectile adj3 dysfunction).tw,kw. (37632)
- 16 (libido adj3 (low\$ or decreas\$ or reduc\$ or loss)).tw,kw. (4556)
- 17 (impotence or impotent).tw,kw. (14431)
- 18 hypogonad\$.tw,kw. (28111)
- 19 (low\$ adj3 testosterone).tw. (11863)
- 20 (deficien\$ adj3 (androgen or gonad\$ or testosterone)).tw. (8162)
- 21 (insuffic\$ adj3 (androgen or gonad\$ or testosterone)).tw. (957)
- 22 (kallman or klinefetter).tw. (181)
- 23 or/8-22 (140363)
- 24 qualitative research/ (96365)
- 25 qualitative research.tw,kw. (35430)
- 26 (qualitative adj3 method\$.tw. (54722)
- 27 (qualitative method? or qualitative methodology).kw. (2814)
- 28 (qualitative adj3 stud\$.tw. (98260)
- 29 qualitative study.kw. (2649)
- 30 focus groups/ use ppez (25196)
- 31 focus group?.tw,kw. (83881)
- 32 grounded theory/ (6130)
- 33 grounded theory.tw,kw. (21668)
- 34 narrative analys?s.tw,kw. (2198)
- 35 process evaluation.tw,kw. (6010)
- 36 mixed method?.tw,kw. (31069)
- 37 mixed method\$.mp. (31948)
- 38 mixed methodology.tw,kw. (723)
- 39 (in depth adj4 interview\$.tw. (42446)

40 in depth interview?.kw. (201)
 41 ((semi structured or semistructured) adj5 interview\$.tw. (91899)
 42 semi structured interview?.kw. (288)
 43 qualitative interview\$.tw. (18260)
 44 qualitative interview?.kw. (443)
 45 (interview\$ and theme\$.tw. (62693)
 46 interview?.kw. (6929)
 47 (interview\$ and audio recorded).tw. (5373)
 48 qualitative case stud\$.tw. (2045)
 49 descriptive case stud\$.tw. (496)
 50 qualitative case study.kw. (25)
 51 descriptive case study.kw. (0)
 52 qualitative exploration.tw,kw. (2043)
 53 qualitative evaluation.tw,kw. (6659)
 54 qualitative intervention.tw,kw. (25)
 55 qualitative approach.tw,kw. (8103)
 56 qualitative inquiry.tw,kw. (1217)
 57 qualitativ\$ analys\$.tw. (33285)
 58 qualitative analysis.kw. (1290)
 59 (qualitative adj3 data).tw. (35636)
 60 qualitative data.kw. (158)
 61 discourse analysis.tw,kw. (3432)
 62 discursive.tw,kw. (3340)
 63 phenomenological.tw,kw. (30865)
 64 thematic analysis.tw,kw. (27748)
 65 ethnograph\$.tw. (19034)
 66 ethnography.kw. (1888)
 67 action research.tw,kw. (7823)
 68 ethno?methodology.tw,kw. (159)
 69 social construction.tw,kw. (1775)
 70 or/24-69 (440293)
 71 phenomenological characteristics.tw,kw. (260)
 72 phenomenological model.tw,kw. (1894)
 73 action research arm test.tw,kw. (1086)
 74 protocol.ti. (83352)
 75 or/71-74 (86550)
 76 70 not 75 (432530)
 77 7 and 76 (236)
 78 23 and 76 (1287)
 79 77 or 78 (1465)
 80 exp animals/ not human/ (8451782)
 81 exp nonhuman/ not humans/ (4820545)
 82 79 not (80 or 81) (1418)
 83 82 and male/ (942)
 84 82 not ((women not men) or (female not male)).tw. (1135)
 85 83 or 84 (1202)
 86 limit 85 to yr="1992-Current" (1156)

Economics review

Database: Embase <1980 to 2018 Week 36>, Ovid MEDLINE® and Epub Ahead of Print, In-process & Other Non-indexed Citations and Daily <1946 to August 31, 2018>

Date of Search: 4 September 2018

- 1 exp androgens/tu use ppez (7644)
- 2 hormone replacement therapy/ use ppez (9276)
- 3 2 and (men or androgen? or testosterone).af. (2599)
- 4 Androgen Therapy/ use emez (5233)
- 5 (androgen replacement therapy or art).tw,kw. (186472)
- 6 testosterone.tw,kw. (161096)
- 7 or/1,3-6 (353641)
- 8 exp Erectile Dysfunction/ use ppez (17852)
- 9 exp impotence/ use emez (38986)
- 10 Sexual Dysfunction, Physiological/ (22084)
- 11 testosterone/df (1227)
- 12 Libido/ use ppez (4539)
- 13 Libido Disorder/ use emez (5707)
- 14 Hypogonadism/ (21676)
- 15 (erectile adj3 dysfunction).tw,kw. (37615)
- 16 (libido adj3 (low\$ or decreas\$ or reduc\$ or loss)).tw,kw. (4557)
- 17 (impotence or impotent).tw,kw. (14430)
- 18 hypogonad\$.tw,kw. (28105)
- 19 (low\$ adj3 testosterone).tw. (11861)
- 20 (deficien\$ adj3 (androgen or gonad\$ or testosterone)).tw. (8163)
- 21 (insuffic\$ adj3 (androgen or gonad\$ or testosterone)).tw. (957)
- 22 (kallman or klinefetter).tw. (181)
- 23 or/8-22 (140335)
- 24 7 and 23 (30346)
- 25 exp "costs and cost analysis"/ use ppez (218103)
- 26 exp economic evaluation/ use emez (273359)
- 27 economics/ (248641)
- 28 health economics/ use emez (27121)
- 29 exp health care cost/ use emez (261474)
- 30 exp economics,hospital/ use ppez (23064)
- 31 exp economics,medical/ use ppez (14042)
- 32 economics,pharmaceutical/ use ppez (2797)
- 33 pharmacoeconomics/ use emez (6785)
- 34 exp models, economic/ use ppez (13505)
- 35 exp decision theory/ (12757)
- 36 monte carlo method/ (59367)
- 37 markov chains/ (15801)
- 38 exp technology assessment, biomedical/ (23409)
- 39 (cost\$ adj2 (effective\$ or utilit\$ or benefit\$ or minimis\$)).ab. (292898)
- 40 economics model\$.tw. (131)
- 41 (economic\$ or pharmacoeconomic\$).tw. (507592)
- 42 (price or prices or pricing).tw. (79068)
- 43 budget\$.tw. (58136)
- 44 (value adj1 money).tw. (59)
- 45 (expenditure\$ not energy).tw. (59417)
- 46 markov\$.tw. (45830)
- 47 monte carlo.tw. (83026)
- 48 (decision\$ adj2 (tree? or analy\$ or model\$)).tw. (45944)
- 49 or/25-48 (1683829)
- 50 (metabolic adj cost).tw. (2520)
- 51 ((energy or oxygen) adj (cost or expenditure)).tw. (56404)

- 52 49 not (50 or 51) (1682066)
- 53 (letter or editorial or note or comment).pt. (3931031)
- 54 52 not 53 (1542896)
- 55 24 and 54 (406)
- 56 exp androgens/ec use ppez (34)
- 57 hypogonadism/ec (8)
- 58 55 or 56 or 57 (430)
- 59 58 not ((women not men) or (female not male)).tw. (395)
- 60 58 and male/ (288)
- 61 59 or 60 (410)
- 62 limit 61 to English language (382)
- 63 limit 62 to yr="1992-Current" (365)

Appendix 2 Data extraction form

TABLE 31 Details of items collected from the included studies

Data extraction section		Information provided in each section								
Study characteristics 1	Linked reports	Publication status	Country	Setting	No. of sites	Study dates	Inclusion criteria	Exclusion criteria	Recruitment method	Allocation method
	Duration of treatment									
Study characteristics 2	Enrolled, <i>n</i>	Randomised, <i>n</i>	Analysed, <i>n</i>	Lost to follow-up, <i>n</i>	Lost to follow-up, reasons	Statistical analysis	Funding source	Conflicts of interest		
Testosterone assessment	Threshold specified in inclusion criteria	Assessment as reported in study								
Intervention characteristics	Testosterone formulation	Dose	Frequency	Route of admin	Other comments	Comparator				
Participant characteristics	Age (years), mean (SD)	BMI (kg/m ²), mean (SD)	Waist circumference (cm), mean (SD)	Current smoker, <i>n</i> (%)	Alcohol use (no of drinks/week), mean (SD)	Diabetes, <i>n</i> (%)	Diabetes duration (year), mean (SD)	Insulin treatment, <i>n</i> (%)	Other comorbidities, <i>n</i> (%)	TT (nmol/l or ng/dl), mean (SD)
	Estradiol (pg/ml), mean (SD)	LH (mIU/ml), mean (SD)	SHBG (nmol/l), mean (SD)	FSH (mIU/ml), mean (SD)	Free T (pg/ml), mean (SD)	HbA1c (%), mean (SD)	Insulin (pmol/l), mean (SD)	HOMA-IR, mean (SD)	Total cholesterol (mmol/l), mean (SD)	LDL (mmol/l), mean (SD)
	HDL (mmol/l), mean (SD)	Triglycerides (mmol/l), mean (SD)	Antidiabetics, <i>n</i> (%)	Statins, <i>n</i> (%)	Antihypertensives, <i>n</i> (%)	Alpha blocking agents, <i>n</i> (%)	5-alpha reductase inhibitors, <i>n</i> (%)	Antidepressants, <i>n</i> (%)	Phosphodiesterase inhibitors, <i>n</i> (%)	PSA (µg/l), mean (SD)
	SBP (mmHg), mean (SD)	DBP (mmHg), mean (SD)	Erectile function, mean (SD)	Lean body mass (kg), mean (SD)	Total fat mass (kg), mean (SD)	Fasting glucose, mg/dl	Hb (g/dl), mean (SD)	Haematocrit (%), mean (SD)	Bone mineral density, g/cm ² , mean (SD)	

Data extraction section	Information provided in each section									
Study characteristics 1	Linked reports	Publication status	Country	Setting	No. of sites	Study dates	Inclusion criteria	Exclusion criteria	Recruitment method	Allocation method
Outcomes	Sexual function	Physical parameter	Functional activities	Psychological symptoms	CV and CBV events	Other comorbidities	Prostate-related outcomes	Physiological markers	QoL	Mortality
Risk of bias	Random sequence generation (selection bias) (low/high/unclear)	Rationale	Allocation concealment (selection bias) (low/high/unclear)	Rationale	Blinding of participants (low/high/unclear)	Rationale	Blinding of personnel (low/high/unclear)	Rationale	Blinding of outcome assessment (detection bias) (low/high/unclear)	Rationale
	Incomplete outcome data (attrition bias) (low/high/unclear)	Rationale	Selective outcome reporting? (reporting bias) (low/high/unclear)	Rationale	Other bias (low/high/unclear)	Rationale				

Appendix 3 List of key items requested from authors of existing trials

TABLE 32 List of key items requested from trials' authors

Type of data	
Study level data	Geographical location (country or countries) in which the trial was carried out)
	Number of trial centres (e.g. a single-centre trial = 1; a multicentre trial will be > 1)
	Number randomised to the TRT group
	Number randomised to the placebo/standard care group
	Setting (primary care, hospital, community)
	Date first patient randomised
	Date final patient randomised
	Date of final patient follow-up
	Inclusion criteria: testosterone (total and/or free) threshold
	Inclusion criteria: all others
	Exclusion criteria:
	Testosterone assay methodology including internal quality control (IQA) and/or external quality control (EQA). IQA/EQA are surveillance systems to ensure assay alignment, for example comparison of standard sample between operators (IQA) or labs (EQA).
	Evidence of testosterone assay performance against Mass Spectrometry (Immunoassay only)
	Details of TRT during the protocol (product, dosing regimen, duration of treatment, etc.)
	Details of comparator (dose, duration of treatment, etc.)
Did the study measure QoL (state tool that was used, e.g. questionnaire, interview)?	
IPD	Baseline characteristics
Patient ID	
Centre ID	
Demography	Age (unit)
	Weight (unit)
	Height (unit)
	Ethnic group
	Date of entry into study/date of randomisation
	Allocated to TRT or placebo

continued

TABLE 32 List of key items requested from trials' authors (*continued*)

Type of data	
<i>Medical history</i>	<p>Previous myocardial infarction or angina</p> <p>Previous stroke</p> <p>History or family history of prostate cancer</p> <p>Glucose, HbA1c or diagnosis of diabetes mellitus (date of diagnosis, any treatment for diabetes)</p> <p>History of atrial fibrillation</p> <p>History of coronary artery disease or bypass graft surgery</p> <p>History of hypertension</p> <p>History of heart failure</p> <p>Evidence of atherosclerosis</p> <p>Any other cardiac comorbidity</p> <p>Lipid measurements or treatment/diagnosis of hyperlipidaemia</p> <p>Sexual symptoms (e.g. spontaneous erections, diagnosis of erectile dysfunction, libido)</p> <p>Physical parameters (e.g. muscle mass and strength, exercise tolerance, body weight, body mass index, total lean body mass, fat mass).</p> <p>Fatigue (please specify any validated score if used)</p> <p>Mood symptoms, for example low mood, depression, anxiety (please specify any validated score if used)</p> <p>Sleep disturbances</p> <p>Cognitive impairment, for example memory loss, dementia (please specify any validated score if used)</p> <p>History of anaemia</p> <p>History of osteoporosis or fracture</p> <p>History of frailty or falls</p> <p>Other – Please include any other baseline characteristics not mentioned</p>
IPD	<p>Outcomes</p> <p>Randomised to control or TRT?</p> <p>Mortality and cause of mortality</p> <p>Sexual function (measured by the IIEF or other validated tools)</p> <p>Prostate-related outcomes (e.g. PSA levels, prostate volume, increase in the International Prostate Symptoms Score)</p> <p>Cardiac outcomes (e.g. CV and CBV events such as myocardial infarction, angioplasty, coronary artery bypass, arrhythmias, peripheral oedema, elevated blood pressure, stroke; incidence of diabetes)</p> <p>Other adverse outcomes: diagnosis of diabetes, hyperlipidaemia, osteopenia/osteoporosis</p> <p>Physiological markers (e.g. blood pressure, Hb concentration, haematocrit; total serum lipid profile, plasma glucose or HbA1c, bone mineral density)</p>

TABLE 32 List of key items requested from trials' authors (*continued*)

Type of data
Sexual symptoms (e.g. spontaneous erections, diagnosis of erectile dysfunction, libido)
Physical parameters (e.g. muscle mass and strength, exercise tolerance, body weight, body mass index, total lean body mass, fat mass).
Psychological symptoms (e.g. cognition by validated score)
Mood outcomes (e.g. diagnosis of depression, psychiatric illness, mood scores)
Functional activities (e.g. running, walking, kneeling; quantified where possible by validated scores such as the SF-36)
QoL [e.g. EQ-5D, HADS, BDI, Epworth Sleepiness Scale (ESS), AMS]
Other
Drop-outs
Date of study discontinuation
Reason for study discontinuation

Appendix 4 Standard operating procedure for management of IPD

Testosterone Effects and Safety in men with low testosterone levels (TestES): an evidence synthesis and economic evaluation

Standard operating procedure

1. PURPOSE

To provide information and instruction to staff on the secure storage and management of data involved in the TESTES NIHR HTA project. Adherence to this SOP is mandatory. Any queries should be directed to Moira Cruickshank (mcruickshank@abdn.ac.uk).

2. ROLES

Role	Current post holder	Contact details
Gateway Manager	Moira Cruickshank	mcruickshank@abdn.ac.uk 01224 438412
Deputy Gateway Manager (will assume role in absence of above)	Miriam Brazzelli	m.brazzelli@abdn.ac.uk 01224 438082

3. RESPONSIBILITIES

3.1. The Gateway Manager is responsible for:

- 3.1.1. Requesting the setting up of a secure file storage area with appropriate permissions from the University of Aberdeen IT Services.
- 3.1.2. Approving and implementing any subsequent changes to permissions.
- 3.1.3. Receiving sensitive data from TESTES study collaborators and ensuring these are properly stored in the secure file storage area. This will be the only live, complete copy of the data.
- 3.1.4. Ensuring that requests for access are valid and that individuals are granted access only to areas for which there is a necessity of access.
- 3.1.5. Ensuring that all data received from TESTES project collaborators are deleted from their device immediately after transfer to the secure file storage area.
- 3.1.6. Ensure that no data in the secure file storage area are transferred to a personal device at any stage.
- 3.1.7. Deleting data from the secure file storage area as appropriate/needed/requested.

3.2. In the absence of the Gateway Manager, the Deputy Gateway Manager is responsible for:

- 3.2.1. Assuming Gateway Manager responsibilities for the data.

4. PROCEDURES

The following actions **MUST** be carried out, and in the order specified:

4.1. Receiving data

- 4.1.1. The Gateway Manager will download the data provided by the TESTES project collaborator.
- 4.1.2. The Gateway Manager will move the data directly into the secure file storage area.
- 4.1.3. The Gateway Manager will access the data in the new location to ensure successful transfer.
- 4.1.4. The Gateway Manager will then delete the copy of the data that is on their device (in the Downloads file).
- 4.1.5. The Gateway Manager will then empty their Recycle Bin (right click on Recycle Bin and click Empty).

4.2. Storage

- 4.2.1. The Gateway Manager will open the file in the secure file storage area and check the data.
- 4.2.2. The Gateway Manager will manage and update all permissions to the data.
- 4.2.3. Other users may only work on the data in the secure file storage area or in other secure areas of the University of Aberdeen network, and must not copy data onto any personal device(s).

4.3. Backup

- 4.3.1. Data will be subject to the standard University of Aberdeen backup schedule, as implemented by IT Services.

4.4. Data deletion

- 4.4.1. On receipt of a request from study collaborator(s) for data deletion, the Gateway Manager will delete the relevant data from the secure file storage area.
- 4.4.2. The Gateway Manager is responsible for advising the relevant party or parties that the deletion request has been completed.

Appendix 5 Study characteristics and participant characteristics of IPD and non-IPD studies

TABLE 33 Study characteristics of the 17 studies that provided IPD

Study details	Study characteristics	Intervention characteristics and testosterone assay
<p>First author, year: Amory 2004 Secondary reports: Page 2005, Vaughan 2007 Country: USA Language: English Publication type: Full text Study name: N/A</p>	<p>Study design: RCT Study setting: Community (recruitment by advertising and direct mailings) No. of centres: 1 Recruitment period: 22-06-1993 to 20-02-1995 Treatment duration: 36 months Main inclusion criteria: TT < 12.1 nmol/l (350 ng/dl) Main exclusion criteria: Severe illness; Paget's disease; smoking or heavy alcohol use; sleep apnoea; haematocrit > 48%; total cholesterol > 300 mg/dl; abnormal kidney, liver, thyroid, adrenal, or pituitary function; regular exercise; prostate issues; urinary postvoid residual > 149 ml; or abnormal transrectal ultrasound</p>	<p>Interventions: A: Testosterone enanthate B: Placebo Route/dose/frequency: A: 200 mg IM every 2 weeks B: 1 ml sesame oil placebo IM every 2 weeks Other information: Study also included a group randomised to testosterone + finasteride, which was excluded from this review. Thus, the testosterone and placebo groups also involved taking a placebo pill daily Testosterone assay: Testosterone was measured using fluoroimmunoassays. The intra-assay and interassay CVs for midrange measurements were 4.5% and 9.5%, respectively</p>
<p>First author, year: Basaria 2010 Secondary reports: Huang 2013, Storer 2014, 2016, Gagliano-Juca 2018 Country: USA Language: English Publication type: Full text Study name: TOM</p>	<p>Study design: RCT Study setting: Boston University Medical Centre, New England Research Institutes and the Veterans Affairs Boston Healthcare System No. of centres: 3 Recruitment period: September 2005 to December 2009 Treatment duration: 6 months Main inclusion criteria: Men ≥ 65 years, TT 100–350 ng/dl or free testosterone < 50 pg/ml, mobility limited Main exclusion criteria: Active cancers, AUA score > 21, SBP or DBP > 160 or > 100 mmHg, respectively, unstable angina, recent MI, untreated severe obstructive sleep apnoea, elevated PSA, alanine or aspartate aminotransferase, creatinine, HbA1c, or haematocrit, BMI > 40 kg/m², congestive heart failure, mobility-limiting disease</p>	<p>Interventions: A: Testosterone gel B: Placebo Route/dose/frequency: A: 10 g of gel containing 100 mg of testosterone, transdermal application once daily B: Placebo gel, identical in appearance, transdermal application once daily Other information: 2 weeks after randomisation, the dose was adjusted if the average of two testosterone measurements was < 500 ng/dl (17.4 nmol/l), in which case the dose was increased to 15 g daily, or > 1000 ng/dl (34.7 nmol/l), in which case the dose was decreased to 5 g daily Testosterone assay: Liquid chromatography tandem mass spectrometry</p>
<p>First author, year: Basaria 2015 Secondary reports: Huang 2016a, 2016c, Storer 2017, Huang 2018, Traustadottir 2018 Country: USA Language: English Publication type: Full text Study name: TEAAM</p>	<p>Study design: RCT Study setting: Charles Drew University, LA, CA, USA; Boston University Medical Centre, Brigham and Women's Hospital, Boston, MA, USA; Kronos Longevity Research Institute, Phoenix, AZ, USA No. of centres: 3 Recruitment period: 01-09-2004 to 12-02-2009 Treatment duration: 3 years Main inclusion criteria: Men ≥ 60 years, TT levels between 100 and 400 ng/dl or free testosterone < 50 pg/ml Main exclusion criteria: Diseases of testes, pituitary or hypothalamus; prostate or breast cancer; severe lower urinary tract symptoms; elevated PSA, alanine aminotransferase and aspartate aminotransferase, creatinine; haemoglobin A1c or haematocrit; unstable angina; heart failure; MI within last 6 months; SBP > 160 mm, Hg, DBP > 100 mm Hg; or BMI > 35</p>	<p>Interventions: A: Testosterone gel B: Placebo Route/dose/frequency: A: 7.5 g of 1% testosterone gel, transdermal application daily B: Placebo gel, transdermal application daily Other information: 2 weeks after randomisation, TT levels were measured 2–12 hours after gel application. If TT concentration was < 500 ng/dl, the testosterone dose was increased to 10 g or if > 900 ng/dl, reduced to 5 g daily. At the same time, the placebo dose was adjusted for another participant in the placebo group by an unblinded observer to maintain blinding Testosterone assay: TT was measured at Quest Diagnostics, San Juan Capistrano, CA, USA, using a Bayer Advia Centaur immunoassay (Siemens Healthcare Diagnostics) after extraction of serum with ethyl acetate and hexane followed by celite chromatography; this assay, validated against liquid chromatography coupled to tandem mass spectrometry, has a sensitivity of 10 ng/dl</p>

TABLE 33 Study characteristics of the 17 studies that provided IPD (continued)

Study details	Study characteristics	Intervention characteristics and testosterone assay
<p>First author, year: Brock 2016</p> <p>Secondary reports: Brock 2015, Maggi 2016</p> <p>Country: Argentina, Canada, Germany, Spain, Great Britain, Italy, South Korea, Puerto Rico, USA</p> <p>Language: English</p> <p>Publication type: Full text</p> <p>Study name: N/A</p>	<p>Study design: RCT</p> <p>Study setting: NR</p> <p>No. of centres: 98</p> <p>Recruitment period: NR</p> <p>Treatment duration: 12 weeks (after 4 weeks screening)</p> <p>Main inclusion criteria: Males \geq 18 years, TT < 300 ng/dl and at least one symptom of testosterone deficiency</p> <p>Main exclusion criteria: HbA1c > 11%, BMI > 37 kg/m², haematocrit \geq 50%, active cancer, or PSA \geq 4 ng/ml</p>	<p>Interventions:</p> <p>A: Topical testosterone</p> <p>B: Placebo</p> <p>Route/dose/frequency:</p> <p>A: 60 mg topical testosterone solution 2% once daily</p> <p>B: Placebo solution, topical application once daily</p> <p>Other information: To maintain blinding, participants were required to apply a dose of study drug from each of four bottles to the axillae each day. A dose adjustment algorithm was used at weeks 4 and 8 based on a single TT level measurement at the preceding visit using an interactive voice response system to maintain blinding. If required, the dose was decreased to 30 mg or increased in 30 mg increments up to a maximum of 120 mg daily</p> <p>Testosterone assay: Analysis was performed at a central laboratory using the liquid chromatography-mass spectrometry/mass spectrometry method</p>
<p>First author, year: Emmelot-Vonk 2008</p> <p>Secondary reports: Nakhai-Pour 2007, Emmelot-Vonk 2009</p> <p>Country: Netherlands</p> <p>Language: English</p> <p>Publication type: Full text</p> <p>Study name: N/A</p>	<p>Study design: RCT</p> <p>Study setting: University medical centre</p> <p>No. of centres: 1</p> <p>Recruitment period: January 2004 to April 2005</p> <p>Treatment duration: 6 months</p> <p>Main inclusion criteria: Testosterone < 13.7 nmol/l and aged 60–80 years.</p> <p>Main exclusion criteria: Recent MI or CBV accident; heart failure; serious liver or renal diseases; haematological abnormalities, epilepsy, migraine, diabetes mellitus, elevated fasting glucose or PSA</p>	<p>Interventions:</p> <p>A: TU capsules</p> <p>B: Placebo</p> <p>Route/dose/frequency:</p> <p>A: Two capsules 40 mg TU twice daily (total dose 160 mg/day)</p> <p>B: Matching placebo; two capsules twice/day</p> <p>Other information: Adherence monitored by capsule counting at each visit</p> <p>Testosterone assay: The serum levels of testosterone and sex hormone-binding globulin were measured with a solid-phase, competitive, chemiluminescent enzyme immunoassay (Immulite 2000, Diagnostic Products Corporation, Los Angeles, CA, USA) at baseline and at the end of the study. The levels of free testosterone and bioavailable testosterone were calculated from TT, sex hormone-binding globulin and albumin concentrations</p>
<p>First author, year: Gianatti 2014a</p> <p>Secondary reports: Gianatti 2014b, 2016</p> <p>Country: Australia</p> <p>Language: English</p> <p>Publication type: Full text</p> <p>Study name: N/A</p>	<p>Study design: RCT</p> <p>Study setting: Tertiary referral centre</p> <p>No. of centres: 1</p> <p>Recruitment period: November 2009 to February 2013</p> <p>Treatment duration: 30 weeks</p> <p>Main inclusion criteria: Men aged 35–70 years with T2D, TT \leq 12.0 nmol/l (346 ng/dl)</p> <p>Main exclusion criteria: Recent testosterone treatment, pituitary or testicular disorder, TT < 5.0 nmol/l (144 ng/dl), elevated PSA level, haematocrit > 0.50, untreated obstructive sleep apnoea, active malignancy, weight > 135 kg, or HbA1c > 8.5% (69 mmol/mol)</p>	<p>Interventions:</p> <p>A: TU injection</p> <p>B: Placebo</p> <p>Route/dose/frequency:</p> <p>A: 1000 mg intramuscular injection at 0, 6, 18 and 30 weeks</p> <p>B: Visually identical placebo injection at 0, 6, 18 and 30 weeks</p> <p>Other information: Injected into upper outer quadrant of buttock</p> <p>Testosterone assay: TT was measured by ECLIA. Although TT was measured by both ECLIA and LCMS/MS, recruitment was based on ECLIA because the LCMS/MS assay was not available for routine clinical use. Therefore, samples were batched and measured by LCMS/MS at study end</p>

continued

TABLE 33 Study characteristics of the 17 studies that provided IPD (continued)

Study details	Study characteristics	Intervention characteristics and testosterone assay
<p>First author, year: Giltay 2010a Secondary reports: Giltay 2009, 2010b, 2010c, Kalinchenko 2010a, 2010b, 2010c, Mskhalaya 2010, Saad 2010a, 2010b, 2010c, 2010d, Tishova 2010a, 2010c Country: Russia Language: English Publication type: Full text Study name: Moscow</p>	<p>Study design: RCT Study setting: Department of Andrology and Urology, Moscow No. of centres: 1 Recruitment period: October 2005 to October 2008 Treatment duration: 30 weeks Main inclusion criteria: Men aged 35–70 years, TT below 12.0 nmol/l or calculated free testosterone level below 225 pmol/l, diagnosis of the metabolic syndrome Main exclusion criteria: prostate cancer, breast cancer; hepatic tumours; hepatic disease; kidney disease with renal failure; abnormal biochemical or haematological laboratory values</p>	<p>Interventions: A: TU injection B: Placebo Route/dose/frequency: A: 1000 mg IM injection at 0, 6 and 18 weeks B: Visually identical placebo, containing castor oil and benzyl benzoate at 0, 6 and 18 weeks Other information: N/A Testosterone assay: Endocrine measurements (i.e. TT, sex hormone-binding protein, and luteinising hormone) were assessed using a Vitros 3600 system (Ortho-Clinical Diagnostics, Johnson and Johnson company, New Brunswick, NJ, USA) with a chemiluminescence immunoassay technology. Free testosterone levels were estimated using the Vermeulen formula</p>
<p>First author, year: Groti 2018 Secondary reports: N/A Country: Slovenia Language: English Publication type: Full text Study name: N/A</p>	<p>Study design: RCT Study setting: Diabetic outpatient clinic No. of centres: 1 Recruitment period: January 2014 to March 2018 Treatment duration: 12 months Main inclusion criteria: Confirmed untreated late-onset hypogonadism, TT < 11 nmol/l and/or FT < 220 pmol/l, men aged > 35 years, BMI ≥ 30 kg/m², type 2 diabetes Main exclusion criteria: Previously treated hypogonadism, insulin therapy, prostate or breast cancer, severe BPH or elevated PSA, severe obstructive sleep apnoea</p>	<p>Interventions: A: TU injection B: Placebo Route/dose/frequency: A: 1000 mg IM injection at first visit, 6 weeks later and then at 10-week intervals B: Placebo injection at first visit, 6 weeks later and then at 10-week intervals Other information: N/A Testosterone assay: TT levels were measured by coated tube RIA (DiaSorin S. p. A., Salluggia, Italy, and Diagnostic Products Corporation, Los Angeles, CA, USA). Calculated free testosterone and bioavailable testosterone levels were derived from the Vermeulen method</p>
<p>First author, year: Hackett 2013 Secondary reports: Hackett 2011a, 2011b, 2012a, 2012b, 2014a, 2014b, 2016, 2017a, 2017b Country: UK Language: English Publication type: Full text Study name: BLAST</p>	<p>Study design: RCT Study setting: General practice No. of centres: 8 Recruitment period: September 2008 to June 2011 Treatment duration: 18 weeks Main inclusion criteria: Men aged ≥ 18, T2D, TT < 12 nmol/l, symptomatic based on AMS Main exclusion criteria: History of testosterone replacement or prostate, breast or hepatic cancer; abnormal DRE; elevated PSA or haematocrit</p>	<p>Interventions: A: TU injection B: Placebo injection Route/dose/frequency: A: 1000 mg IM injection at weeks 0, 6 and 18 B: Visually identical placebo, containing castor oil and benzyl benzoate at 0, 6 and 18 weeks Other information: Injection administered over 5 minutes into upper outer buttock Testosterone assay: TT was measured using a Roche common platform immunoassay (validated against mass spectrometry)</p>

TABLE 33 Study characteristics of the 17 studies that provided IPD (continued)

Study details	Study characteristics	Intervention characteristics and testosterone assay
<p>First author, year: Hildreth 2013 Secondary reports: N/A Country: USA Language: English Publication type: Full text Study name: N/A</p>	<p>Study design: RCT Study setting: Community No. of centres: 1 Recruitment period: 6 January 2005 to 6 March 2009 Treatment duration: 12 months Main inclusion criteria: Men \geq 60 years old, two total T samples between 200 and 350 ng/dl; BMI $<$ 35 kg/m² Main exclusion criteria: Active coronary artery disease, abnormal DRE, elevated PSA or haematocrit, history of prostate or breast cancer, diabetes, untreated dyslipidaemia</p>	<p>Interventions: A: Testosterone gel B: Placebo gel Route/dose/frequency: A: Daily transdermal application B: Daily transdermal application Other information: The T gel and matching placebo were provided in 2.5- or 5.0-g packets. All subjects were initiated on two 2.5-g packets daily (two placebo packets in the placebo group, one T gel and one placebo packet in the lower-range T group, and two Tgel packets in the higher-range T group). Serum T levels were monitored by the pharmacist approximately every 2 weeks for the first 12 weeks with dose titrations made in 2.5-g increments to achieve a level of 400–550 ng/dl in the lower-range group and 600–1000 ng/dl in the higher-range group; sham adjustments were made in the placebo group. The maximum dose of Tgel used was 10 g/day Testosterone assay: Total serum T and SHBG were measured by ELISA using a Beckman Coulter (Brea, CA) Access II analyser</p>
<p>First author, year: Ho 2012 Secondary reports: Tong 2010, Tan 2011, Tong 2012, Tan 2013 Country: Malaysia Language: English Publication type: Full text Study name: N/A</p>	<p>Study design: RCT Study setting: Tertiary medical centre No. of centres: 1 Recruitment period: May 2008 to February 2010 Treatment duration: 42 weeks Main inclusion criteria: TT \leq 12 nmol/l; symptoms of hypogonadism and PSA $<$ 4 ng/ml Main exclusion criteria: Uncontrolled diabetes mellitus; hypothyroidism/hyperthyroidism; haematocrit $>$ 55%, known prostate cancer; androgen-dependent carcinoma of the male mammary gland; past or existing liver tumours; testosterone treatment in last 6 months</p>	<p>Interventions: A: TU injection B: Placebo Route/dose/frequency: A: 1000 mg TU IM injection at weeks 0, 6, 18, 30 and 42 B: Identical placebo injection at weeks 0, 6, 18, 30 and 42 Other information: Injections given at slow bolus IM at the gluteal region over 1 minute Testosterone assay: TT was measured by immunoassay using a AxSYM testosterone assay (Abbott Laboratories, Wiesbaden, Germany), based on microparticle enzyme immunoassay technology which was confirmed by liquid chromatography/tandem mass spectrometry. The normal reference range for this assay was between 8 and 35 mmol/l, which is consistent with other laboratories</p>
<p>First author, year: Magnussen 2016 Secondary reports: Botha 2017, Magnussen 2017a, 2017b, 2017c Country: Denmark Language: English Publication type: Full text Study name: N/A</p>	<p>Study design: RCT Study setting: Odense University Hospital No. of centres: 1 Recruitment period: April 2012 to November 2013 Treatment duration: 24 weeks Main inclusion criteria: White men, aged 50–70 years, BioT $<$ 7.3 nmol/l, T2D, and receiving metformin for $>$ 3 months Main exclusion criteria: BMI \geq 40 kg/m², elevated HbA1c, haematocrit, or PSA, clinically significant disease of the heart, lung or kidneys, known malignant disease, severe hypertension</p>	<p>Interventions: A: Testosterone gel B: Placebo Route/dose/frequency: A: 5 g testosterone gel daily transdermal application B: Placebo gel daily transdermal application Other information: Testosterone levels were evaluated after 3 weeks of treatment and the dose was increased to 10 g gel daily if BioT level was $<$ 7.3 nmol/l. Compliance was monitored at weeks 3, 12 and 24 concerning gel application, timing, cutaneous area and adverse skin reactions</p>

continued

TABLE 33 Study characteristics of the 17 studies that provided IPD (continued)

Study details	Study characteristics	Intervention characteristics and testosterone assay
<p>First author, year: Marks 2006 Secondary reports: N/A Country: USA Language: English Publication type: Full text Study name: N/A</p>	<p>Study design: RCT Study setting: Urological Sciences Research Centre, Los Angeles No. of centres: 1 Recruitment period: February 2003 to November 2004 Treatment duration: 6 months Main inclusion criteria: Men aged 44–78 years with symptoms attributable to late-onset hypogonadism and testosterone < 300 ng/dl (< 10.4 nmol/l) Main exclusion criteria: Use in the past 6 months of any drug potentially affecting the pituitarygonadal axis, PSA > 10.0 ng/ml, presence of prostate cancer</p>	<p>Testosterone assay: Testosterone and 17β-estradiol levels were measured between 07:30 and 09:00 hours by liquid chromatography tandem mass spectrometry after ether extraction (Statens Serum Institut, Copenhagen, Denmark). FreeT and BioT levels were calculated, using www.issam.ch/freetesto.htm. A single measurement of testosterone was performed to determine eligibility, with BioT as a determinant for lowered testosterone</p> <p>Interventions: A: Testosterone enanthate injection B: Placebo injection Route/dose/frequency: A: 150 mg testosterone enanthate IM injection biweekly B: Saline placebo IM injection biweekly Other information: Compliance with dosing exceeded 99% – only two doses (one in each group) of testosterone (of 533) were given 'out of window' (i.e. > 3 weeks after the prior dose). In these two participants, a new 2-week cycle was established based on the timing of the last dose, but overall study duration did not exceed 28 weeks in either participant Testosterone assay: Determined by mass spectroscopy</p>
<p>First author, year: Merza 2006 Secondary reports: N/A Country: UK Language: English Publication type: Full text Study name: N/A</p>	<p>Study design: RCT Study setting: Sheffield male sexual dysfunction clinic No. of centres: 1 Recruitment period: NR Treatment duration: 6 months Main inclusion criteria: (1) serum total T < 10 nmol/l and/or a free androgen index (FAI) < 30% (total T/SHBG \times 100), (2) absence of known prostate or breast cancer, prostatic hypertrophy, raised PSA (> 2.5 ng/l), uncontrolled hypertension, diabetes mellitus, uncontrolled cardiac disease, renal failure (creatinine > 150 μmol/l), liver disease, polycythaemia (haematocrit > 50%), Exclusion criteria: NR</p>	<p>Interventions: A: Testosterone body patch B: Placebo body patch Route/dose/frequency: A: Testosterone body patch 60 cm² containing 328 mg testosterone, delivering 5 mg/day B: Placebo body patch Other information: All patients were advised to administer one patch every morning at the same time Testosterone assay: TT was measured by IRMA (Orion Diagnostics); the intra-assay CV was 7% at 2.01 nmol/l and 4.8% at 52.7 nmol/l; the inter-assay CV was 7% and 4.8% respectively. Reference range for total T was 8–38 nmol/l</p>

TABLE 33 Study characteristics of the 17 studies that provided IPD (continued)

Study details	Study characteristics	Intervention characteristics and testosterone assay
<p>First author, year: Snyder 2016</p> <p>Secondary reports: Cunningham 2016 (sexual function trial), Budoff 2017 (CV trial), Snyder 2017 (bone trial), Mohler 2018 (CV biomarkers)</p> <p>Country: USA</p> <p>Language: English</p> <p>Publication type: Full text</p> <p>Study name: T-trials</p>	<p>Study design: RCT</p> <p>Study setting: University medical centres</p> <p>No. of centres: 12</p> <p>Recruitment period: June 2010 to June 2013</p> <p>Treatment duration: 12 months</p> <p>Main inclusion criteria: Men \geq 65 years old; total T < 275 ng/dl, one or more symptoms potentially consequent to low T</p> <p>Main exclusion criteria: Diagnosed prostate cancer or high risk of PC, severe lower urinary tract symptoms, untreated sleep apnoea, MI or stroke in previous 3 months or elevated BP</p>	<p>Interventions:</p> <p>A: Testosterone gel</p> <p>B: Placebo gel</p> <p>Route/dose/frequency:</p> <p>A: Testosterone gel initial dose 5 g daily, transdermal application to abdomen, shoulder or upper arms. The serum testosterone concentration was measured monthly for the first 3 months. If the testosterone concentration was not between 500 and 800 ng/dl at any time point, the dose was either increased by increments of 1.25–2.5 g/day, up to a maximum of 15 g/day, or decreased by increments of 1.25–3.75 ng/day. If the serum testosterone concentration was > 800 ng/dl following two consecutive reductions in AndroGel dose, treatment was discontinued. A placebo participant was also discontinued</p> <p>B: Placebo gel daily transdermal application to abdomen, shoulders or upper arms</p> <p>Other information: Serum testosterone concentration was measured at months 1, 2, 3, 6 and 9 in a central laboratory (Quest Clinical Trials), and the dose of testosterone gel was adjusted after each measurement in an attempt to keep the concentration within the normal range for young men (19–40 years of age). To maintain blinding when the dose was adjusted in a participant receiving testosterone, the dose was changed simultaneously in a participant receiving placebo</p> <p>Testosterone assay: At the end of the trials, the serum concentrations of TT, free testosterone, dihydrotestosterone, estradiol and sex hormone-binding globulin were measured in serum samples frozen at -80°C. Steroid assays were performed at the Brigham Research Assay Core Laboratory (Boston) by liquid chromatography with tandem mass spectroscopy, and free testosterone was measured by equilibrium dialysis. All samples from each participant were measured in the same assay run</p>
<p>First author, year: Srinivas-Shankar 2010</p> <p>Secondary reports: Atkinson 2010</p> <p>Country: UK</p> <p>Language: English</p> <p>Publication type: Full text</p> <p>Study name: N/A</p>	<p>Study design: RCT</p> <p>Study setting: Community</p> <p>No. of centres: 1</p> <p>Recruitment period: NR</p> <p>Treatment duration: 6 months</p> <p>Main inclusion criteria: Community-dwelling men aged \geq 65 years, at least one frailty criterion and total T < 12 nmol/l (345 ng/dl)</p> <p>Main exclusion criteria: Prostate cancer, benign prostatic hyperplasia, PSA > 4 ng/ml, chronic renal impairment, active liver disease, moderate to severe peripheral vascular disease, severe chronic obstructive pulmonary disease, congestive heart failure, untreated sleep apnoea</p>	<p>Interventions:</p> <p>A: Testosterone gel</p> <p>B: Placebo gel</p> <p>Route/dose/frequency:</p> <p>A: Testosterone gel 50 mg once daily transdermal application</p> <p>B: Placebo gel once daily transdermal application</p> <p>Other information: The dose of gel was adjusted to 75 or 25 mg/day according to serum T at day 10 and 3 months. Dose adjustment was undertaken if T levels remained outside the target range (18–30 nmol/l)</p>

continued

TABLE 33 Study characteristics of the 17 studies that provided IPD (continued)

Study details	Study characteristics	Intervention characteristics and testosterone assay
<p>First author, year: Svartberg 2008</p> <p>Secondary reports: Agledahl 2008, Agledahl 2009</p> <p>Country: Norway</p> <p>Language: English</p> <p>Publication type: Full text</p> <p>Study name: TROMSØ study</p>	<p>Study design: RCT</p> <p>Study setting: Clinical research unit of the University Hospital of North Norway</p> <p>No. of centres: 1</p> <p>Recruitment period: 2005</p> <p>Treatment duration: 40 weeks</p> <p>Main inclusion criteria: Men aged 60–80 years with serum testosterone \leq 11.0 nmol/l</p> <p>Main exclusion criteria: Prostate cancer or other malignancies, unstable ischaemic or congestive heart disease, epilepsy, migraine, elevated haematocrit, Hb, PSA > 4.0, creatinine or alanine aminotransferase > 100 U/l</p>	<p>Testosterone assay: Levels of testosterone were measured by chemiluminescent immunoassay with a Roche Elecys E170 platform at baseline, 10 days and 3 and 6 months</p> <p>Interventions: A: TU injection B: Placebo injection</p> <p>Route/dose/frequency: A: TU 1000 mg IM depot injection at weeks 0, 6, 16, 28, 40 B: Placebo injection IM at weeks 0, 6, 16, 28, 40</p> <p>Other information: N/A</p> <p>Testosterone assay: Serum TT, follicle-stimulating hormone, luteinising hormone, estradiol and PSA were analysed by ECLIA using an automated clinical chemistry analyser (Modular E; Roche Diagnostics GmbH, Mannheim, Germany). The total analytical precision expressed as the sum of intra- and inter-assay coefficients of variation (CVa) were 5.6%, 4.0%, 2.2%, 4.7% and 3.2%, respectively</p>

BioT, bioavailable testosterone; CV, coefficient of variation; DRE, digital rectal examination; ECLIA, electrochemiluminescence immunoassay; ELISA, enzyme-linked immunosorbent assay; IM, intramuscular; LCMS/MS, liquid chromatography-tandem mass spectroscopy; MEIA, microparticle enzyme immunoassay; N/A, not applicable; NR, not reported; SHBG, sex hormone-binding globulin; T, testosterone.

TABLE 34 Study characteristics of the 18 studies that did not provide IPD

Study details	Study characteristics	Intervention characteristics
<p>First author, year: Aversa 2010a</p> <p>Secondary reports: N/A</p> <p>Country: Italy</p> <p>Language: English</p> <p>Publication type: Full text</p> <p>Study name: N/A</p>	<p>Study design: RCT</p> <p>Study setting: NR</p> <p>No. of centres: 1</p> <p>Recruitment period: NR</p> <p>Treatment duration: 12 months (plus 12 months open-label extension, not included in this review)</p> <p>Main inclusion criteria: 45–65 years of age, MS and/or T2DM and total T below 3.0 ng/ml (11 nmol/l) or calculated free t < 250 pmol/l (10 pg/ml), and \geq2 symptoms of hypogonadism</p> <p>Main exclusion criteria: Prostate or breast cancer; history of tumours; symptomatic obstructive sleep apnoea; elevated haematocrit or PSA, abnormal DRE; any uncontrolled endocrine disorder; heart failure; hepatic or renal insufficiency</p>	<p>Interventions: A: TU injection plus placebo gel B: Placebo injection plus placebo gel</p> <p>Route/dose/frequency: A: TU IM injection every 12 weeks from week 6 plus placebo gel transdermal application daily B: Placebo IM injection every 12 weeks from week 6 plus placebo gel transdermal application daily</p> <p>Other information: To ensure that treatment remained blinded, a standard double-dummy technique was used. All patients received ampoule for injections, either TU or a visually identical placebo injection and gel sachets containing placebo preparations</p> <p>Testosterone assay: TT measured by electrochemiluminescence (method Immulite 2000 Siemens, Milan, Italy; within and between-assay coefficients of variation were 5.1% and 7.2%)</p>

TABLE 34 Study characteristics of the 18 studies that did not provide IPD (continued)

Study details	Study characteristics	Intervention characteristics
<p>First author, year: Aversa 2010b</p> <p>Secondary reports: Aversa 2010c</p> <p>Country: Italy</p> <p>Language: English</p> <p>Publication type: Full text</p> <p>Study name: N/A</p>	<p>Study design: RCT</p> <p>Study setting: NR</p> <p>No. of centres: NR</p> <p>Recruitment period: NR</p> <p>Treatment duration: 12 months</p> <p>Main inclusion criteria: 50–65 years old with metabolic syndrome and/or T2DM, TT < 3.20 ng/ml (11 nmol/l), and ≥2 symptoms of hypogonadism</p> <p>Main exclusion criteria: Prostate or breast cancer; history of tumours; obstructive sleep apnoea; elevated PSA or haematocrit, abnormal DRE; uncontrolled diabetes and/or treatment with insulin; severe cardiac, hepatic or renal insufficiency</p>	<p>Interventions:</p> <p>A: TU capsules</p> <p>B: TU injection</p> <p>C: Placebo</p> <p>Route/dose/frequency:</p> <p>A: TU two capsules of 40 mg twice/day (crossed over at 6 months to IM TU)</p> <p>B: TU IM injections 1000 mg every 12 weeks from week 6</p> <p>C: Placebo gel preparation 3–4 g/day</p> <p>Other information: All IM injections were administered into the gluteus muscle by the same trained physician</p> <p>Testosterone assay: TT measured by electrochemiluminescence (method Immulite 2000 Siemens, Milan, Italy; within and between-assay coefficients of variation were 5.1% and 7.2%)</p>
<p>First author, year: Basurto 2008</p> <p>Secondary reports: N/A</p> <p>Country: Mexico</p> <p>Language: English</p> <p>Publication type: Full text</p> <p>Study name: N/A</p>	<p>Study design: RCT</p> <p>Study setting: General Hospital 25 of the Instituto Mexicano del Seguro Social, Mexico City</p> <p>No. of centres: 1</p> <p>Recruitment period: November 2002 – January 2005</p> <p>Treatment duration: 12 months</p> <p>Main inclusion criteria: Men ≥ 60 years, TT between 100 and 400 ng/dl</p> <p>Main exclusion criteria: Testes, pituitary, or hypothalamus disease; cancers; severe lower urinary tract symptoms, elevated PSA, alanine aminotransferase/aspartate aminotransferase, creatinine, HbA1c or haematocrit; unstable angina; heart failure; MI in last 6 months; elevated BP; or BMI > 35</p>	<p>Interventions:</p> <p>A: Testosterone enanthate</p> <p>B: Placebo</p> <p>Route/dose/frequency:</p> <p>A: Testosterone enanthate 250 mg IM injection every 21 days for 12 months</p> <p>B: Placebo IM injection every 21 days for 12 months</p> <p>Other information: N/A</p> <p>Testosterone assay: Serum testosterone levels were measured by a specific solid-phase radioimmunoassay using commercial kits from Diagnostic Products Corporation (Los Angeles, CA, USA). The intra- and inter-assay CVs were 7.3% and 7% respectively. The analytical sensitivity of this assay was 4 ng/dl</p>
<p>First author, year: Behre 2012</p> <p>Secondary reports: N/A</p> <p>Country: Austria, Finland, Germany, Ireland, Italy, Spain, Sweden, UK</p> <p>Language: English</p> <p>Publication type: Full text</p> <p>Study name: N/A</p>	<p>Study design: RCT</p> <p>Study setting: NR</p> <p>No. of centres: NR</p> <p>Recruitment period: NR</p> <p>Treatment duration: 6 months (plus 12 months open-label extension, not included in this review)</p> <p>Main inclusion criteria: Men 50–80 years with symptoms of testosterone deficiency including TT < 15 nmol/l</p> <p>Main exclusion criteria: BMI > 35 kg/m²; prostate or breast cancer; elevated PSA, haematocrit, or prolactin; sleep apnoea, polycythaemia, hypothalamic pituitary disorders, uncontrolled diabetes mellitus, thyroid disorders, hypertension or epilepsy; severe cardiac and hepatic or renal insufficiency</p>	<p>Interventions:</p> <p>A: Hydroalcoholic testosterone gel</p> <p>B: Placebo</p> <p>Route/dose/frequency:</p> <p>A: 5 g hydroalcoholic 1% testosterone gel daily transdermal application</p> <p>B: Placebo gel daily transdermal application</p> <p>Other information: At month 3 (visit 4) there was an option to increase the dose based on the clinical response of the patient. Depending on the clinical response of the patient, and if TT was not above 28 nmol/l (> 8.1 ng/ml), the dose of testosterone or placebo gel could be increased after 3 months to 7.5 g (75 mg testosterone)</p> <p>Testosterone assay: Levels of serum TT and bioavailable testosterone were assessed at a central laboratory. The testosterone levels were measured by ECLIA technique on a Roche Elecsys or Modular E170 analyzer. The intra-assay CV was 2.1% and inter-assay CV was 2.8% for the range of testosterone values measured. Bioavailable testosterone was calculated using the formula of Vermeulen</p>

continued

TABLE 34 Study characteristics of the 18 studies that did not provide IPD (continued)

Study details	Study characteristics	Intervention characteristics
<p>First author, year: Borst 2014 Secondary reports: Beggs 2014 Country: USA Language: English Publication type: Full text Study name: N/A</p>	<p>Study design: RCT Study setting: NR No. of centres: NR Recruitment period: NR Treatment duration: 12 months Main inclusion criteria: Men aged ≥ 60, TT ≤ 300 ng/dl Main exclusion criteria: Failure of the Mini-Cog test, prostate or breast cancer, severe BPH, AUA/IPSS score ≥ 25, congestive heart failure, sleep apnoea, elevated Hct or PSA, BMI > 35, orthopaedic limitations</p>	<p>Interventions: A: Testosterone enanthate B: Placebo Route/dose/frequency: A: Testosterone enanthate weekly IM injection 125 mg/week B: Placebo IM injection weekly Other information: Study design was (finasteride OR placebo) AND (testosterone OR vehicle). Only participants in the testosterone and placebo groups were included in this review Testosterone assay: Testosterone was assayed by Cobas electrochemiluminescence immunoassay</p>
<p>First author, year: Cavallini 2004 Secondary reports: N/A Country: Italy Language: English Publication type: Full text Study name: N/A</p>	<p>Study design: RCT Study setting: NR No. of centres: NR Recruitment period: 02-01-2002 to 30-06-2002 Treatment duration: 6 months Main inclusion criteria: ≥ 60 years old with symptoms of androgen decline Main exclusion criteria: Lower urinary tract obstructive symptoms, prostate volume > 20 cm³, increased PSA, suspicion of cancer on prostate DRE, recent MI or major surgery, diabetes, untreated hypertension, cardiovascular disease</p>	<p>Interventions: A: TU B: Placebo Route/dose/frequency: A: TU 160 mg/day oral application B: Placebo tablet daily Other information: Third arm of study: 45 patients, mean age 66 years, range 61–73, used propionyl-L-carnitine 2 g/day plus acetyl-L-carnitine 2 g/day (not included in this review) Testosterone assay: TT was measured by recombinant immunoassay after extraction and celite chromatography. The intra- and inter-assay coefficients of variation were 6% and 13.5%, respectively. The assay does not cross-react with methyl-testosterone. Free testosterone was calculated as the product of the total and percentage of dialysable free testosterone. 15 The intra- and inter-assay coefficients of variation were 5% and 6.6%, respectively. The reagents were from Diagnostic Products (Los Angeles, CA, USA)</p>
<p>First author, year: Cherrier 2015 Secondary reports: N/A Country: USA Language: English Publication type: Full text Study name: N/A</p>	<p>Study design: RCT Study setting: The Veterans Affairs Puget Sound Health Care System (VAPSHCS) in Seattle, WA, USA No. of centres: 1 Recruitment period: NR Treatment duration: 6 months Main inclusion criteria: Aged 60–90 years; diagnosis of MCI T < 300 ng/dl; (4) AUA score ≤ 19; and BMI < 35 Main exclusion criteria: haematocrit $> 50\%$, severe symptoms of BPH, significant liver/renal/heart/peripheral vascular/pulmonary disease, insulin-dependent diabetes mellitus, obstructive sleep apnoea, prostate or breast cancer; blood pressure $> 160/90$</p>	<p>Interventions: A: Testosterone gel B: Placebo Route/dose/frequency: A: Testosterone gel 50–100 mg daily transdermal application B: Placebo gel 50–100 mg transdermal application Other information: Target TT level of 500–900 ng/dl Testosterone assay: Testosterone was measured by liquid chromatography tandem mass spectrometry and free testosterone calculated using formula of Vermeulen. Mean intra- and inter-assay coefficients of variation are 4.9% and 7.1% for TT</p>

TABLE 34 Study characteristics of the 18 studies that did not provide IPD (continued)

Study details	Study characteristics	Intervention characteristics
<p>First author, year: Chiang 2007 Secondary reports: Chiang 2009 Country: Taiwan Language: English Publication type: Full text Study name: N/A</p>	<p>Study design: RCT Study setting: Two hospitals in Taipei No. of centres: 2 Recruitment period: November 2002 to November 2004 Treatment duration: 3 months Main inclusion criteria: Men aged 20–75, diagnosed with testosterone deficiency requiring testosterone replacement Main exclusion criteria: elevated PSA or haematocrit, pulmonary, renal, hepatic, neurological, musculoskeletal or cardiovascular disease, sleep apnoea, BMI > 27.0 or < 18.5</p>	<p>Interventions: A: Testosterone gel B: Placebo gel Route/dose/frequency: A: Testosterone gel 50 mg daily transdermal application B: Placebo gel daily transdermal application Other information: Participants were instructed to apply the gel once a day to two application sites, either shoulders and upper arms and/or the abdomen. Alternative application sites were continued throughout the study. The gel was not to be applied to the genitals. If a participant developed skin irritation at the application site, he was advised to apply corticosteroid cream before applying the gel Testosterone assay: Laboratory examination of PSA and other hormone test profiles was performed using a radioimmunoassay kit</p>
<p>First author, year: Clague 1999 Secondary reports: N/A Country: UK Language: English Publication type: Full text Study name: N/A</p>	<p>Study design: RCT Study setting: Withington Hospital, Manchester No. of centres: 1 Recruitment period: NR Treatment duration: 12 weeks Main inclusion criteria: age > 60 years, TT < 14 nmol/l Exclusion criteria: NR</p>	<p>Interventions: A: Testosterone enanthate B: Placebo Route/dose/frequency: A: Testosterone enanthate 200 mg IM injection at 2 weekly intervals B: Placebo IM injection at 2 weekly intervals Other information: N/A Testosterone assay: NR</p>
<p>First author, year: Dhindsa 2016a Secondary reports: Dhindsa 2016b, 2016c, Ghanim 2018 Country: USA Language: English Publication type: Full text Study name: N/A</p>	<p>Study design: RCT Study setting: Division of Endocrinology, Diabetes and Metabolism, State University of New York at Buffalo No. of centres: 1 Recruitment period: December 2010 to January 2014 Treatment duration: 22 weeks Main inclusion criteria: Males with type 2 diabetes, 30–65 years old, HbA1c ≤ 8% (64 mmol/mol), and stable diabetes regimen for 3 months Main exclusion criteria: Use of androgens, glucocorticoids, or opiates in last 6 months, pan-hypopituitarism, congenital HH, prolactinoma, severe hepatic or kidney disease, PSA > 4 ng/ml</p>	<p>Interventions: A: Testosterone cypionate B: Placebo Route/dose/frequency: A: Testosterone cypionate 250 mg every 2 weeks IM injection in the buttock B: Placebo injection IM every 2 weeks in the buttock Other information: Dose of testosterone was adjusted to keep cFT concentrations in normal range (6.5–25 ng/dl) Testosterone assay: TT and estradiol concentrations were measured by liquid chromatography–tandem mass spectrometry. Tracer equilibrium dialysis is considered the gold standard for measuring free steroid hormone concentrations, and this methodology was used to separate the free testosterone and free estradiol (Nichols Institute, Chantilly, VA, USA, and San Juan Capistrano, CA, USA)</p>

continued

TABLE 34 Study characteristics of the 18 studies that did not provide IPD (continued)

Study details	Study characteristics	Intervention characteristics
<p>First author, year: Dias 2016</p> <p>Secondary reports: Dias 2014, Dias 2015, Dias 2017a, Dias 2017b</p> <p>Country: USA</p> <p>Language: English</p> <p>Publication type: Full text</p> <p>Study name: N/A</p>	<p>Study design: RCT</p> <p>Study setting: National Institute on Ageing</p> <p>No. of centres: 1</p> <p>Recruitment period: March 2004 to January 2015</p> <p>Treatment duration: 12 months</p> <p>Main inclusion criteria: Men aged ≥ 65 years, total T < 350 ng/dl</p> <p>Main exclusion criteria: haematocrit < 36%, MMSE score < 24, polycythaemia, osteoporosis, history of stroke or diabetes, uncontrolled high blood pressure, severe BPH recent acute coronary syndrome, use of bisphosphonate, selective oestrogen receptor modulator or any anabolic agents</p>	<p>Interventions:</p> <p>A: Testosterone gel</p> <p>B: Placebo</p> <p>Route/dose/frequency:</p> <p>A: Testosterone gel 5 g daily transdermal application</p> <p>B: Placebo gel daily</p> <p>Other information: Study involved 3 groups: T gel + placebo tablet; Anastrozole tablet + placebo gel; placebo gel + placebo tablet. Only the testosterone and placebo groups were included in this review</p> <p>Testosterone assay: Testosterone levels were measured by liquid chromatography tandem mass spectroscopy. Bioavailable testosterone was measured using ammonium sulphate precipitation method. Testosterone detection limit was 2.5 ng/dl and intra-assay CV was 3.5%; inter-assay CV 5.3%; bioavailable testosterone detection limit was 4.7 mg/dl; intra-assay CV 2.0%; inter-assay CV 2.3%</p>
<p>First author, year: Jones 2011</p> <p>Secondary reports: Buvat 2009, Jones 2009, 2010, Stanworth 2014</p> <p>Country: Belgium, France, Germany, Italy, Netherlands, Spain, Sweden, UK</p> <p>Language: English</p> <p>Publication type: Full text</p> <p>Study name: TIMES2</p>	<p>Study design: RCT</p> <p>Study setting: Outpatient centres</p> <p>No. of centres: 36</p> <p>Recruitment period: February 2006 to March 2007</p> <p>Treatment duration: 12 months</p> <p>Main inclusion criteria: Men aged ≥ 40 years, confirmed hypogonadism with ≥ 2 symptoms and type 2 diabetes and/or metabolic syndrome</p> <p>Main exclusion criteria: Recent TRT or insulin therapy; prostate or breast cancer; abnormal DRE; severe symptomatic BPH, elevated PSA</p>	<p>Interventions:</p> <p>A: Testosterone gel</p> <p>B: Placebo gel</p> <p>Route/dose/frequency:</p> <p>A: 3 g metered-dose 2% testosterone gel once daily transdermal application</p> <p>B: Placebo gel once daily</p> <p>Other information: Treatment was applied daily (07:00–10:00) to clean, dry, intact skin on the thighs or abdomen. TT was measured at 2, 4 and 12 weeks with dose adjustments made as follows: TT > 52 nmol/l, testosterone dose reduced to 40 mg/day; TT < 17 nmol/l, dose increased to 80 mg/day. Dummy dose changes were performed in the placebo group to maintain blinding</p> <p>Testosterone assay: Free testosterone was calculated from testosterone, albumin, and sex hormone-binding globulin, using the Vermeulen equation</p>
<p>First author, year: Kaufman 2011a</p> <p>Secondary reports: Kaufman 2011b, 2011c, Morgentaler 2012, 2014</p> <p>Country: USA</p> <p>Language: English</p> <p>Publication type: Full text</p> <p>Study name: N/A</p>	<p>Study design: RCT</p> <p>Study setting: Clinics</p> <p>No. of centres: 63</p> <p>Recruitment period: February 2007 to April 2007</p> <p>Treatment duration: 182 days</p> <p>Main inclusion criteria: Hypogonadal men 18–80 years old, serum TT < 300 and BMI ≥ 18 kg/m² to ≤ 40 kg/m²</p> <p>Main exclusion criteria: impaired liver function, IPSS-1 > 15, PSA > 2.5 ng/ml, or abnormal DRE, prostate or breast cancer, sleep apnoea, heart failure, and haematocrit > 48%, or Hb > 16 g/dl</p>	<p>Interventions:</p> <p>A: Testosterone gel</p> <p>B: Placebo</p> <p>Route/dose/frequency:</p> <p>A: Testosterone 1.62% gel, titrated doses of 1.25 g, 2.5 g, 3.75 g or 5 g, with all men started on 2.5 g, daily transdermal application</p> <p>B: Placebo gel, daily transdermal application</p> <p>Other information: All eligible subjects were started at a dose of 2.5 g 1.62% testosterone gel or matching placebo on day 1 of the study. Subjects returned to the clinic at day 14, day 28, and day 42 for pre-dose serum TT assessments and other secondary assessments. Within 2 days of these visits, the subject's dose was titrated up or down in 1.25 g increments if TT levels were not within the prespecified range of 350–750 ng/dl. No dose was titrated below 1.25 g or above 5.0 g during the study. Sham titrations occurred in placebo-treated subjects to maintain blinding. Subjects were maintained at their respective day 42 dose until day 182</p>

TABLE 34 Study characteristics of the 18 studies that did not provide IPD (continued)

Study details	Study characteristics	Intervention characteristics
<p>First author, year: Kenny 2010 Secondary reports: N/A Country: USA Language: English Publication type: Full text Study name: N/A</p>	<p>Study design: RCT Study setting: Major medical institution No. of centres: 1 Recruitment period: NR Treatment duration: 12 months Main inclusion criteria: Men with testosterone < 350 ng/dl; osteoporosis and frailty Main exclusion criteria: PSA > 6.5 ng/dl, prostate cancer; bone metabolism or; pituitary disease; sleep apnoea; metastatic or advanced cancer; advanced liver or renal disease; and Hb > 16.5 g/dl</p>	<p>Testosterone assay: Serum concentrations of testosterone were assayed by Pharmaceutical Product Development (PPD, Inc., Richmond, VA, USA) using validated liquid chromatography/tandem mass spectrometry methodology. The analyte measured in each case refers to all forms of testosterone present in the serum including free and reversibly protein-bound species. Serum TT concentrations used for screening and titration were measured by immunoassay by Quintiles Laboratories, Smyrna, GA, USA</p> <p>Interventions: A: Testosterone gel B: Placebo gel Route/dose/frequency: A: Testosterone 1% gel, 5 mg daily, transdermal application B: Placebo gel daily transdermal Other information: Adherence assessment included measurement of returned gel bottles and monthly application logs. All men were counselled to maintain calcium intake of 1500 mg/day and received Citracal (Bayer Healthcare, LLC, Morristown, NJ, USA) 315-mg tablets to meet these goals; most men consumed three to four tablets per day to supplement their diet. In addition, all men were given 1000 IU of cholecalciferol per day</p> <p>Testosterone assay: Total and bioavailable testosterone and sex-hormone-binding globulin measurements were performed at Endocrine Sciences Inc., Calabasas Hills, CA, USA. Testosterone levels were measured by radioimmunoassay, and bioavailable testosterone by competitive binding of the non-SHBG-bound portion of testosterone following ammonium sulphate precipitation of the SHBG-bound steroid as described by Nankin. Intra-assay variability of the testosterone assay is < 7%, bioavailable testosterone < 4%</p>
<p>First author, year: Morales 2009 Secondary reports: N/A Country: Canada Language: English Publication type: Full text Study name: N/A</p>	<p>Study design: RCT Study setting: Ambulatory clinics No. of centres: 4 Recruitment period: NR Treatment duration: 4 months Main inclusion criteria: Clinical picture compatible with TDS; total T levels < 12 nmol/l and DHEAS of < 3.5 µmol/l Main exclusion criteria: Contraindication for use of androgens, including history of prostate cancer, PSA > 4 ng/l, or a history of substance abuse within 2 years</p>	<p>Interventions: A: TU capsules B: Placebo capsules Route/dose/frequency: A: TU capsules 80 mg, twice daily, oral application B: Placebo capsules, twice daily, oral application Other information: Patients were advised to take the medication with meals, which is a fundamental requirement for absorption of oral TU Testosterone assay: NR</p>

continued

TABLE 34 Study characteristics of the 18 studies that did not provide IPD (continued)

Study details	Study characteristics	Intervention characteristics
<p>First author, year: Paduch 2015a Secondary reports: Paduch 2015b Country: USA, Canada, Mexico Language: English Publication type: Full text Study name: N/A</p>	<p>Study design: RCT Study setting: Medical centres No. of centres: NR Recruitment period: August 2011 to December 2013 Treatment duration: 16 weeks Main inclusion criteria: Men \geq 26 years of age, total T levels $<$ 300 ng/dl ($<$ 10.41 nmol/l) and \geq 1 ejaculation symptom Main exclusion criteria: Ejaculation or erectile dysfunction, prostate or breast cancer, haematocrit \geq 50%, significant lower urinary tract symptoms and BMI $>$ 35 kg/m²</p>	<p>Interventions: A: Testosterone solution B: Placebo solution Route/dose/frequency: A: Testosterone 2% solution 60 mg daily transdermal application B: Placebo solution daily transdermal application Other information: Applied daily to the axillae for 16 weeks with the goal of maintaining serum total T levels in the range 300–1050 ng/dl (10.41–36.44 nmol/l). 4 weeks after randomisation, participants with serum T $<$ 300 ng/dl ($<$ 10.41 nmol/l) were titrated up to 90 mg daily, whereas those with serum T $>$ 1050 ng/dl ($>$ 36.44 nmol/l) were titrated down to 30 mg daily. Treatment compliance was monitored by weighing study medication bottles at each clinic visit. Participants were considered compliant if at least 70% of their expected daily doses were taken Testosterone assay: Testosterone measurements were performed using LC-MS/MS in a central laboratory enrolled in the Centers for Disease Control and Prevention hormone harmonisation programme</p>
<p>First author, year: Steidle 2003 Secondary reports: Seftel 2003 Country: USA Language: English Publication type: Full text Study name: N/A</p>	<p>Study design: RCT Study setting: Clinics No. of centres: 43 Recruitment period: NR Treatment duration: 90 days Main inclusion criteria: 20–80 years of age and T level of \leq 10.4 nmol/l and \geq 1 symptom of low T Main exclusion criteria: use of oestrogen therapy, LHRH antagonist, human GH therapy; or history of drug abuse within 12 months</p>	<p>Interventions: A: Testosterone gel 50 mg B: Testosterone gel 100 mg C: Testosterone patch D: Placebo gel Route/dose/frequency: A: Testosterone gel 50 mg daily transdermal application B: Testosterone gel 100 mg daily transdermal application C: Two testosterone patches \times 2.5 mg (each containing 12.2 mg testosterone), daily transdermal application D: Placebo gel daily transdermal application Other information: The testosterone and placebo gel were identical and applied as two tubes of 50 mg T (100 mg/day), one tube of 50 mg T and one tube of placebo (50 mg/day) or two tubes of placebo. All study drug treatments were applied in the morning; repeat applications occurred at the same time of day for the duration of the study. Each day in the gel-treated group, patients applied the contents of two tubes. The content of one tube was applied to one shoulder and the content of the remaining tube was applied to the other shoulder. Patients allocated to receive the T patch applied two adhesive patches daily. Application sites included the back, abdomen, upper arms and thighs. Patches were to be worn for 24 hours and then replaced each morning at approximately the same time. Subjects randomised to 1 of the 2 T gel arms could be titrated at day 60 on the basis of their day 30 T pharmacokinetic profile. Subjects were titrated from 50 to 100 mg/d at day 60 if their day 30 mean serum T concentration (Cavg) was $<$ 300 ng/dl (10.4 nmol/l). Subjects were titrated from 100 to 50 mg/d at day 60 if their day 30 T Cavg was $>$ 1000 ng/dl (34.7 nmol/l). These titration decisions were undertaken by a third-party physician who was unaware of any clinical aspects of the individual subjects Testosterone assay: Serum testosterone levels were measured at ICON Laboratories (Farmingdale, NY, USA), using validated radioimmunoassay kits</p>

TABLE 34 Study characteristics of the 18 studies that did not provide IPD (continued)

Study details	Study characteristics	Intervention characteristics
First author, year: Wang 2013 Secondary reports: N/A Country: China Language: English Publication type: Full text Study name: N/A	Study design: RCT Study setting: NR No. of centres: 1 Recruitment period: NR Treatment duration: 24 months Main inclusion criteria: Men (aged > 60 years) with osteoporosis and serum T < 300 ng/dl Main exclusion criteria: Prostate tumour; cancer; poorly controlled diabetes, uncontrolled hypertension; hypothyroidism or hyperthyroidism; hyperparathyroidism; abnormal liver function or renal disease	Interventions: A: Testosterone capsules 40 mg B: Testosterone capsules 20 mg C: Placebo capsules Route/dose/frequency: A: Testosterone capsules 40 mg daily oral application B: Testosterone capsules 20 mg daily oral application C: Placebo capsules daily oral application Other information: All patients were also supplemented with calcium (600 mg) and vitamin D3 (125 IU) daily. Participants were requested to maintain their habitual diet and exercise patterns Testosterone assay: Serum concentrations of TT, free testosterone and estradiol were analysed by chemical luminescence method.

AUA, American Urological Association; BP, blood pressure; BPH, benign prostatic hyperplasia; cFT, calculated free testosterone; CV, coefficient of variation, DHEAS, dehydroepiandrosterone sulphate; DRE, digital rectal examination; ECG, electrocardiogram; ECLIA, electrochemiluminescence immunoassay; FT, free testosterone, GH, growth hormone; IM, intramuscular, IPSS, International Prostate Symptom Score; LHRH, luteinising hormone-releasing hormone; N/A, not applicable; NR, not reported; Hct, haematocrit; T, testosterone; WHO, World Health Organization.

TABLE 35 Participant characteristics of the 17 studies that provided IPD

Study ID, interventions (n randomised)	Participant characteristics [all mean (SD) unless otherwise specified]	
First author, year: Amory 2004 Interventions: A: TRT IM (n = 24) B: Placebo (n = 24)	Age, years: A: 71 (4), B: 71 (5) BMI, kg/m²: A: 28.7 (3.6), B: 27.9 (3.6) TT, nmol/l: A: 9.9 (1.6), B: 10.5 (1.7) Free testosterone, pg/ml: A: NR, B: NR HbA1c, %: A: NR, B: NR Total cholesterol, mmol/l: A: NR, B: NR LDL, mmol/l: A: NR, B: NR HDL, mmol/l: A: NR, B: NR	Triglycerides, mmol/l: A: NR, B: NR Haematocrit, %: A: NR, B: NR Hb, g/dl: A: NR, B: NR Diabetes, n (%): A: NR, B: NR Other comorbidities, n (%): A: NR, B: NR SBP, mmHg: A: NR, B: NR DBP, mmHg: A: NR, B: NR PSA, ng/dl: A: 0.9 (0.8), B: 1.4 (1.1)
First author, year: Basaria 2010 Interventions: A: TRT gel (n = 106) B: Placebo (n = 103)	Age, years: A: 74 (6), B: 74 (5) BMI, kg/m²: A: 29.7 (4.1), B: 30.0 (4.2) TT, nmol/l: A: 250 (57) ng/dl, B: 236 (66) ng/dl Free testosterone, pg/ml: A: 48 (12), B: 43 (14) HbA1c, %: A: NR, B: NR Total cholesterol, mmol/l: A: 165 (35), B: 171 (39) LDL, mmol/l: A: 89 (30), B: 92 (33) HDL, mmol/l: A: 46 (13), B: 48 (18)	Triglycerides, mmol/l: A: 159 (111), B: 143 (69) Haematocrit, %: A: NR, B: NR Hb, g/dl: A: NR, B: NR Diabetes, n (%): A: NR, B: NR Other comorbidities, n (%): A: Pre-existing CV disease: 56 (52.8%); obesity: 48 (45.3%); hypertension: 90 (84.9%); hyperlipidaemia: 67 (63.2%) B: Pre-existing CV disease: 48 (46.6%); obesity: 50 (48.9%); hypertension: 80 (77.7%); hyperlipidaemia: 51 (49.5%) SBP, mmHg: A: 137 (15), B: 137 (14) DBP, mmHg: A: 77 (10), B: 75 (10) PSA, ng/dl: A: NR, B: NR

continued

TABLE 35 Participant characteristics of the 17 studies that provided IPD (continued)

Study ID, interventions (n randomised)	Participant characteristics [all mean (SD) unless otherwise specified]	
First author, year: Basaria 2015 Interventions: A: TRT gel (n = 156) B: Placebo (n = 152)	Age, years: A: 66.9 (5.0), B: 68.3 (5.3) BMI, kg/m²: A: 28.1 (2.1), B: 28.0 (2.9) TT, nmol/l: A: 307.2 (64.3) ng/dl, B: 307.4 (67.4) ng/dl Free testosterone, pg/ml: A: 64.0 (17.2), B: 60.9 (18.0) HbA1c, %: A: 5.7 (0.8), B: 5.7 (0.7) Total cholesterol, mmol/l: A: 187.2 (42.1), B: 183.4 (36.7) LDL, mmol/l: A: 115.6 (35.2), B: 109.7 (31.9) HDL, mmol/l: A: 47.1 (12), B: 48.7 (14.2)	Triglycerides, mmol/l: A: 142.7 (87.9), B: 138.9 (76.4) Haematocrit, %: A: 43.7 (3.7), B: 43.6 (3.6) Hb, g/dl: A: 14.6 (1.2), B: 14.4 (1.6) Diabetes, n (%): A: 22 (14.2), B: 24 (15.9) Other comorbidities, n (%): A: Obesity: 40 (25.5%), hypertension: 71 (45.8%), hyperlipidaemia: 80 (51.6%); prior coronary artery disease: 24 (15.5%) B: Obesity: 42 (27.8%), hypertension: 57 (37.7%), hyperlipidaemia: 77 (51.0%); prior coronary artery disease: 22 (14.6%) SBP, mmHg: A: NR, B: NR DBP, mmHg: A: NR, B: NR PSA, ng/dl: A: 1.31 (0.82), B: 1.25 (0.94)
First author, year: Brock 2016 Interventions: A: TRT solution (n = 358) B: Placebo (n = 357)	Age, years: A: 54.7 (10.6), B: 55.9 (11.4) BMI, kg/m²: A: 30.3 (4.1), B: 30.9 (4.2) TT, nmol/l: A: 202.6 (66.3) ng/dl, B: 201.2 (67.3) ng/dl Free testosterone, pg/ml: A: NR, B: NR HbA1c, %: A: NR, B: NR Total cholesterol, mmol/l: A: NR, B: NR LDL, mmol/l: A: NR, B: NR HDL, mmol/l: A: NR, B: NR	Triglycerides, mmol/l: A: NR, B: NR Haematocrit, %: A: NR, B: NR Hb, g/dl: A: NR, B: NR Diabetes, n (%): A: NR, B: NR Other comorbidities, n (%): A: NR, B: NR SBP, mmHg: A: NR, B: NR DBP, mmHg: A: NR, B: NR PSA, ng/dl: A: NR, B: NR
First author, year: Emmelot-Vonk 2008 Interventions: A: TRT capsules (n = 120) B: Placebo (n = 117)	Age, years: A: 67.1 (5.0), B: 67.4 (4.9) BMI, kg/m²: A: 27.4 (3.8), B: 27.3 (3.9) TT, nmol/l: A: 11.0 (1.9), B: 10.5 (1.8) Free testosterone, pg/ml: A: 0.22 (0.02) nmol/l, B: 0.21 (0) nmol/l HbA1c, %: A: NR, B: NR Total cholesterol, mmol/l: A: NR, B: NR LDL, mmol/l: A: NR, B: NR HDL, mmol/l: A: NR, B: NR	Triglycerides, mmol/l: A: NR, B: NR Haematocrit, %: A: NR, B: NR Hb, g/dl: A: NR, B: NR Diabetes, n (%): A: NR, B: NR Other comorbidities, n (%): A: Previous CV disease, including MI, angina, hypertension and stroke: 35/113 (31.0) B: Previous CV disease, including MI, angina, hypertension and stroke: 37/110 (33.6) SBP, mmHg: A: 155 (23.3), B: 151.4 (22.7) DBP, mmHg: A: 89.2 (12.0), B: 86.8 (11.7) PSA, ng/dl: A: NR, B: NR
First author, year: Gianatti 2014a Interventions: A: TRT IM (n = 45) B: Placebo (n = 43)	Age, years: Median (IQR) A: 62 (58–68), B: 62 (57–67) BMI, kg/m²: Median (IQR): 31.5 (28.3–35.5), B: 33.4 (31.4–35.4) TT, nmol/l: Median (IQR): A: 8.7 (7.1–11.1), B: 8.5 (7.2–11.0) Free testosterone, pg/ml: Median (IQR): 183 (148–247) pmol/l, B: 187 (150–237) pmol/l HbA1c, %: Median (IQR): A: 6.8 (6.4–7.6), B: 7.1 (6.7–7.5) Total cholesterol, mmol/l: Median (IQR): A: 4.2 (3.8–4.8), B: 4.5 (3.6–4.8) LDL, mmol/l: Median (IQR): A: 2.3 (1.7–2.8), B: 2.2 (1.8–2.8) HDL, mmol/l: Median (IQR): A: 1.1 (0.9–1.3), B: 1.0 (0.8–1.2)	Triglycerides, mmol/l: Median (IQR): A: 1.6 (1.1–2.4), B: 1.8 (1.3–2.4) Haematocrit, %: Median (IQR): A: 0.44 (0.41–0.46), B: 0.43 (0.41–0.45) Hb, g/dl: Median (IQR): A: 151 (139–157) g/l, B: 151 (142–156) g/l Diabetes, n (%): A: 45 (100), B: 43 (100) Other comorbidities, n (%): Metabolic syndrome, %: A: 98, B: 95 SBP, mmHg: Median (IQR): A: 140 (130–150), B: 140 (129–150) DBP, mmHg: Median (IQR): A: 72 (70–80), B: 80 (70–82) PSA, ng/dl: Median (IQR): A: 0.84 (0.58–1.24), B: 0.73 (0.46–1.26)

TABLE 35 Participant characteristics of the 17 studies that provided IPD (continued)

Study ID, interventions (n randomised)	Participant characteristics [all mean (SD) unless otherwise specified]	
First author, year: Giltay 2010a Interventions: A: TRT IM (n = 113) B: Placebo (n = 71)	Age, years: Mean (95% CI): A: 51.6 (49.8–53.4), B: 52.8 (50.5–55.0) BMI, kg/m²: Mean (95% CI): A: 35.3 (34.2–36.6), B: 34.2 (32.9–35.7) TT, nmol/l: Mean (95% CI): A: 6.7 (6.0–7.4), B: 7.5 (6.6–8.5) Free testosterone, pg/ml: Mean (95% CI): A: 120 (107–135) pM, B: 130 (113–151) pM HbA1c, %: A: NR, B: NR Total cholesterol, mmol/l: Mean (95% CI): A: 5.6 (5.4–5.8), B: 5.6 (5.3–5.9) LDL, mmol/l: Mean (95% CI): A: 3.7 (3.5–3.9), B: 3.7 (3.4–3.9) HDL, mmol/l: Mean (95% CI): A: 1.13 (1.06–1.2), B: 1.12 (1.03–1.21)	Triglycerides, mmol/l: Mean (95% CI): A: 2.04 (1.84–2.26), B: 2.27 (2.00–2.57) Haematocrit, %: A: NR, B: NR Hb, g/dl: A: NR, B: NR Diabetes, n (%): A: 32 (28.3), B: 24 (33.8) Other comorbidities, n (%): Hypertension: A: 100 (88.5), B: 61 (85.9) SBP, mmHg: A: NR, B: NR DBP, mmHg: A: NR, B: NR PSA, ng/dl: A: NR, B: NR
First author, year: Groth 2018 Interventions: A: TRT IM (n = 28) B: Placebo (n = 27)	Age, years: A: NR, B: NR (overall mean 60.15, SD 7.23, range 40–70) BMI, kg/m²: A: 34.03 (4.37), B: 32.63 (3.67) TT, nmol/l: A: 7.24 (1.97), B: 7.96 (1.34) Free testosterone, pg/ml: A: 155.54 (41.11) pmol/l, B: 192.07 (44.23) pmol/l HbA1c, %: A: 8.12 (1.04), B: 7.89 (0.77) Total cholesterol, mmol/l: A: 5.31 (0.91), B: 5.31 (0.97) LDL, mmol/l: A: 2.79 (0.77), B: 2.80 (0.95) HDL, mmol/l: A: 1.01 (0.22), B: 1.05 (0.32)	Triglycerides, mmol/l: A: 2.86 (1.49), B: 3.52 (3.15) Haematocrit, %: A: NR, B: NR Hb, g/dl: A: NR, B: NR Diabetes, n (%): A: NR, B: NR Other comorbidities, n (%): Lipid abnormalities: A: 28 (100), B: 27 (100) SBP, mmHg: A: 134.64 (10.71), B: 138.15 (13.24) DBP, mmHg: A: 77.50 (5.85), B: 78.89 (5.25) PSA, ng/dl: A: NR, B: NR
First author, year: Hackett 2013 Interventions: A: TRT IM (n = 97) B: Placebo (n = 102)	Age, years: A: 61.2 (10.5), B: 62.0 (9.3) BMI, kg/m²: A: 33.0 (6.1), B: 32.4 (5.5) TT, nmol/l: A: 9.2 (3.5), B: 8.9 (3.5) Free testosterone, pg/ml: A: 187.7 (57.0) pmol/l, B: 181.2 (63.6) pmol/l HbA1c, %: A: 7.74 (1.31), B: 7.47 (1.24) Total cholesterol, mmol/l: A: 4.16 (0.91), B: 4.09 (0.90) LDL, mmol/l: A: 2.21 (0.81), B: 2.19 (0.91) HDL, mmol/l: A: 1.15 (0.65), B: 1.09 (0.33)	Triglycerides, mmol/l: A: 2.0 (1.5), B: 2.0 (1.1) Haematocrit, %: A: NR, B: NR Hb, g/dl: A: NR, B: NR Diabetes, n (%): A: 92 (100), B: 98 (100) Other comorbidities, n (%): A: NR, B: NR SBP, mmHg: A: 140.2 (15.9), B: 137.1 (13.0) DBP, mmHg: A: 79.4 (9.4), B: 77.5 (8.9) PSA, ng/dl: A: 1.4 (1.4), B: 1.4 (1.2)
First author, year: Hildreth 2013 Interventions: A: TRT gel (n = 55) B: Placebo (n = 28)	Age, years: A: 66.4 (5.0), B: 67.5 (5.6) BMI, kg/m²: A: NR, B: NR TT, nmol/l: A: 300.0 (42.4) ng/dl, B: 301.1 (41.4) ng/dl Free testosterone, pg/ml: A: NR, B: NR HbA1c, %: A: NR, B: NR Total cholesterol, mmol/l: A: 174.5 (36.3) mg/dl, B: 181.0 (27.3) mg/dl LDL, mmol/l: A: 100.9 (28.2) mg/dl, B: 109.0 (28.2) mg/dl HDL, mmol/l: A: 42.5 (10.2) mg/dl, B: 45.3 (10.8) mg/dl	Triglycerides, mmol/l: A: 155.2 (80.3) mg/dl, B: 134.0 (54.8) mg/dl Haematocrit, %: A: 46.5 (3.0), B: 46.1 (3.3) Hb, g/dl: A: NR, B: NR Diabetes, n (%): A: NR, B: NR Other comorbidities, n (%): A: NR, B: NR SBP, mmHg: A: 126.2 (15.2), B: 128.1 (18.5) DBP, mmHg: A: 75.6 (7.1), B: 74.1 (9.2) PSA, ng/dl: A: 1.4 (1.0) ng/ml, B: 1.4 (0.8) ng/ml

continued

TABLE 35 Participant characteristics of the 17 studies that provided IPD (continued)

Study ID, interventions (n randomised)	Participant characteristics [all mean (SD) unless otherwise specified]	
First author, year: Ho 2102 Interventions: A: TRT IM (n = 60) B: Placebo (n = 60)	Age, years: A: 53.4 (7.4), B: 53.0 (8.2) BMI, kg/m²: A: 30.4 (5.2), B: 28.2 (4.5) TT, nmol/l: A: 8.9 (2.0), B: 9.1 (1.8) Free testosterone, pg/ml: A: NR, B: NR HbA1c, %: A: NR, B: NR Total cholesterol, mmol/l: A: NR, B: NR LDL, mmol/l: A: NR, B: NR HDL, mmol/l: A: NR, B: NR	Triglycerides, mmol/l: A: NR, B: NR Haematocrit, %: A: NR, B: NR Hb, g/dl: A: NR, B: NR Diabetes, n (%): A: 14 (23.3), B: 9 (15) Other comorbidities, n (%): A: Hypertension: 29 (48.3%), dyslipidaemia: 22 (36.7%); coronary artery disease: 6 (10%) B: Hypertension: 20 (33.3%), dyslipidaemia: 20 (33.3%); coronary artery disease: 3 (5%) SBP, mmHg: A: 132.5 (14.6), B: 127.9 (10.4) DBP, mmHg: A: 84.1 (7.4), 81.4 (6.4) PSA, ng/dl: A: NR, B: NR
First author, year: Magnussen 2016 Interventions: A: TRT gel (n = 22) B: Placebo (n = 21)	Age, years: A: 61 (6), B: 59 (6) BMI, kg/m²: Arithmetic mean (IQR): A: 30.6 (28.9–32.3), B: 30.8 (28.9–32.6) TT, nmol/l: Median (IQR): A: 7.1 (6.6–11.9), B: 9.4 (8.1–12.5) Free testosterone, pg/ml: Median (IQR): A: 0.20 (0.15–0.26) nmol/l, B: 0.24 (0.21–0.28) nmol/l HbA1c, %: Geometric mean (95% CI): A: 6.5 (6.3–6.8), B: 6.5 (6.2–6.8) Total cholesterol, mmol/l: Arithmetic mean (SD): A: 4.0 (0.7), B: 3.8 (1.1) LDL, mmol/l: Geometric mean (95% CI): A: 2.2 (1.9–2.5), B: 2.0 (1.7–2.5) HDL, mmol/l: Geometric mean (95% CI): A: 1.01 (0.90–1.12), B: 0.93 (0.86–1.00)	Triglycerides, mmol/l: Geometric mean (95% CI): A: 1.3 (1.0–1.7), B: 1.5 (1.2–1.8) Haematocrit, %: Arithmetic mean (SD): A: 43.1 (0.02), B: 43.3 (0.02) Hb, g/dl: Geometric mean (95% CI): A: 9.0 (8.7–9.2) mmol/l, B: 9.1 (8.9–9.3) mmol/l Diabetes, n (%): A: 20 (100), B: 19 (100) Other comorbidities, n (%): A: NR, B: NR SBP, mmHg: A: 137.7 (16.7), B: 138.2 (12.8) DBP, mmHg: A: 80.5 (11.1), B: 81.7 (8.2) PSA, ng/dl: Geometric mean (95% CI): A: 0.7 (0.5–1.0), B: 1.0 (0.6–1.4)
First author, year: Marks 2006 Interventions: A: TRT IM (n = 22) B: Placebo (n = 22)	Age, years: Median (range): A: 68 (44–78), B: 70 (45–78) BMI, kg/m²: Median (range): A: 28.34 (22.7–37.9), B: 29.57 (23.6–37.8) TT, nmol/l: Median (range): A: 221 (163–320) ng/dl, B: 252 (144–328) ng/dl Free testosterone, pg/ml: Median (range): A: 48 (17–102), B: 51 (16–66) HbA1c, %: A: NR, B: NR Total cholesterol, mmol/l: A: NR, B: NR LDL, mmol/l: A: NR, B: NR HDL, mmol/l: A: NR, B: NR	Triglycerides, mmol/l: A: NR, B: NR Haematocrit, %: Median (range): A: 43.2 (35.2–50.5), B: 43.6 (37.4–48.2) Hb, g/dl: Median (range): A: 14.5 (11–18), B: 14.9 (12.6–16.1) Diabetes, n (%): A: NR, B: NR Other comorbidities, n (%): A: NR, B: NR SBP, mmHg: A: NR, B: NR DBP, mmHg: A: NR, B: NR PSA, ng/dl: Median (range): A: 1.55 (0.27–5.78) ng/ml, B: 0.97 (0–2.47) ng/ml
First author, year: Merza 2006 Interventions: A: TRT body patch (n = 20) B: Placebo (n = 19)	Age, years: A: 63 (9), B: 59.7 (10.2) BMI, kg/m²: A: NR, B: NR TT, nmol/l: A: 8.4 (3.3), B: 7.5 (2.5) Free testosterone, pg/ml: A: NR, B: NR HbA1c, %: A: NR, B: NR Total cholesterol, mmol/l: A: NR, B: NR LDL, mmol/l: A: NR, B: NR HDL, mmol/l: A: NR, B: NR	Triglycerides, mmol/l: A: NR, B: NR Haematocrit, %: A: NR, B: NR Hb, g/dl: A: NR, B: NR Diabetes, n (%): A: NR, B: NR Other co-morbidities, n (%): A: NR, B: NR SBP, mmHg: A: NR, B: NR DBP, mmHg: A: NR, B: NR PSA, ng/dl: A: NR, B: NR
First author, year: Snyder 2016 Interventions: A: TRT gel (n = 395) B: Placebo (n = 395)	Age, years: A: 72.1 (5.7), B: 72.3 (5.8) BMI, kg/m²: A: 31.0 (3.5), B: 31.0 (3.6) TT, nmol/l: A: 232 (63) ng/dl, B: 236 (67) ng/dl Free testosterone, pg/ml: A: 62 (21.4), B: 65 (23.4) HbA1c, %: A: NR, B: NR Total cholesterol, mmol/l: A: NR, B: NR LDL, mmol/l: A: NR, B: NR HDL, mmol/l: A: NR, B: NR	Triglycerides, mmol/l: A: NR, B: NR Haematocrit, %: A: NR, B: NR Hb, g/dl: A: NR, B: NR Diabetes, n (%): A: 148 (37.5), B: 144 (36.5) Other comorbidities, n (%): A: Hypertension: 286 (72.4); history of MI: 53 (13.4); history of stroke: 16 (4.1); sleep apnoea: 78 (19.8) B: Hypertension: 280 (70.9); history of MI: 63 (16); history of stroke: 17 (4.3); sleep apnoea: 76 (19.2) SBP, mmHg: A: NR, B: NR DBP, mmHg: A: NR, B: NR PSA, ng/dl: A: NR, B: NR

TABLE 35 Participant characteristics of the 17 studies that provided IPD (continued)

Study ID, interventions (n randomised)	Participant characteristics [all mean (SD) unless otherwise specified]	
First author, year: Srinivas-Shankar 2010 Interventions: A: TRT gel (n = 138) B: Placebo (n = 136)	Age, years: A: 73.7 (5.7), B: 73.9 (6.4) BMI, kg/m²: A: 27.9 (4.1), B: 27.7 (4.0) TT, nmol/l: A: 11.0 (3.2), B: 10.9 (3.1) Free testosterone, pg/ml: A: 180 (50) pmol/l, B: 180 (50) pmol/l HbA1c, %: A: NR, B: NR Total cholesterol, mmol/l: Median (IQR): A: 4.6 (3.9–5.3), B: 4.6 (3.9–5.3) LDL, mmol/l: Median (IQR): A: 2.5 (1.7–3.0), B: 2.3 (1.7–2.9) HDL, mmol/l: Median (IQR): A: 1.4 (1.2–1.6), B: 1.5 (1.1–1.8)	Triglycerides, mmol/l: Median (IQR): A: 1.5 (1.0–2.1), B: 1.4 (1.0–2.0) Haematocrit, %: A: 44 (3.0), B: 42 (4.0) Hb, g/dl: A: 14.6 (1.2), B: 14.2 (1.3) Diabetes, n (%): A: NR, B: NR Other comorbidities, n (%): A: NR, B: NR SBP, mmHg: A: NR, B: NR DBP, mmHg: A: NR, B: NR PSA, ng/dl: A: 1.5 (0.9) ng/ml, B: 1.5 (0.9) ng/ml
First author, year: Svartberg 2008 Interventions: A: TRT IM (n = 14) B: Placebo (n = 13)	Age, years: A: 68.9 (5.4), B: 69.3 (5.0) BMI, kg/m²: A: 30.6 (3.9), B: 29.4 (3.9) TT, nmol/l: A: 8.5 (1.7), B: 8.2 (2.4) Free testosterone, pg/ml: A: 183.7 (29.0) pmol/l, B: 183.4 (65.6) pmol/l HbA1c, %: A: NR, B: NR Total cholesterol, mmol/l: A: 5.03 (1.53), B: 5.02 (1.09) LDL, mmol/l: A: 3.15 (0.84), 3.02 (0.81) HDL, mmol/l: A: 1.30 (0.34), B: 1.31 (0.24)	Triglycerides, mmol/l: A: 1.44 (0.79), B: 1.1 (0.52) Haematocrit, %: A: NR, B: NR Hb, g/dl: A: NR, B: NR Diabetes, n (%): A: NR, B: NR Other comorbidities, n (%): A: Coronary vascular disease (acute MI, cerebral stroke, angina pectoris): 3 (23.1), B: Coronary vascular disease (acute MI, cerebral stroke, angina pectoris): 5 (38.5) SBP, mmHg: A: NR, B: NR DBP, mmHg: A: NR, B: NR PSA, ng/dl: A: NR, B: NR

IQR, interquartile range; IM, intramuscular; MI, myocardial infarction; NR, not reported.

TABLE 36 Participant characteristics of the 18 studies that did not provide IPD

Study ID, interventions (n randomised)	Participant characteristics [all mean (SD) unless otherwise specified]	
First author, year: Aversa 2010a Interventions: A: TRT IM (n = 40) B: Placebo (n = 10)	Age, years: A: 58 (10), B: 57 (8) BMI, kg/m²: A: 30.2 (4.5), B: 31 (6.2) TT, nmol/l: A: 8.33 (2.4), B: 9 (1.7) Free testosterone, pg/ml: A: 26 (8.6), B: 27 (7.3) HbA1c, %: A: 5.7 (0.5), B: 6.6 (1.3) Total cholesterol, mmol/l: Median (quartiles), mg/dl: A: 210 (180–240), B: 215 (170–250) LDL, mmol/l: A: NR, B: NR HDL, mmol/l: Median (quartiles), mg/dl: A: 44 (33–55), B: 42 (37–47)	Triglycerides, mmol/l: Median (quartiles), mg/dl: A: 155 (120–190), B: 145 (115–175) Haematocrit, %: A: NR, B: NR Hb, g/dl: A: NR, B: NR Diabetes, n (%): A: NR, B: NR Other comorbidities, n (%): Metabolic syndrome (MS) only: A: 28 (70), B: 7 (70) T2DM + MS: A: 12 (30), B: 3 (30) SBP, mmHg: A: 137 (7), B: 138 (16) DBP, mmHg: A: 83 (9), B: 84 (12) PSA, ng/dl: A: 1.07 (0.4), B: 1.1 (0.5)
First author, year: Aversa 2010b Interventions: A: TRT oral (n = 10) B: TRT IM (n = 32) C: Placebo (n = 10)	Age, years: A: 57 (8), B: 58 (10), 55 (5) BMI, kg/m²: A: 32.5 (5.2), B: 30.2 (4.5), C: 31 (6.2) TT, nmol/l: A: NR, B: NR, C: NR (all groups < 3 ng/ml or 300 ng/dl at baseline) Free testosterone, pg/ml: A: NR, B: NR, C: NR HbA1c, %: A: 5.8 (0.5), B: 5.7 (0.5), C: 6.3 (1.2) Total cholesterol, mmol/l: A: 206 (30) mg/dl, B: 210 (33) mg/dl, C: 217 (51) mg/dl LDL, mmol/l: A: NR, B: NR, C: NR HDL, mmol/l: A: 42 (9) mg/dl, B: 44 (11) mg/dl, C: 146 (32) mg/dl	Triglycerides, mmol/l: A: 150 (34) mg/dl, B: 158 (36) mg/dl, C: 146 (32) mg/dl Haematocrit, %: A: NR, B: NR, C: NR Hb, g/dl: A: NR, B: NR, C: NR Diabetes, n (%): A: 3 (30.0), B: 10 (31.2), C: 4 (40.0) Other comorbidities, n (%): MS only: A: 7 (70), B: 22 (69), C: 6 (60) T2D + MS: A: 3 (30), B: 8 (25), C: 3 (30) SBP, mmHg: A: 140 (10), B: 136 (8), C: 138 (16) DBP, mmHg: A: 82 (14), B: 84 (12), C: 84 (12) PSA, ng/dl: A: NR, B: NR, C: NR

continued

TABLE 36 Participant characteristics of the 18 studies that did not provide IPD (continued)

Study ID, interventions (n randomised)	Participant characteristics [all mean (SD) unless otherwise specified]	
First author, year: Basurto 2008 Interventions: A: TRT IM (n = 25) B: Placebo (n = 23)	Age, years: A: 63.2 (7.9), B: 63.1 (7.7) BMI, kg/m²: A: 27.4 (3.0), B: 27.2 (2.0) TT, nmol/l: A: 301 (32), B: 310 (37) ng/dl Free testosterone, pg/ml: A: NR, B: NR HbA1c, %: A: NR, B: NR Total cholesterol, mmol/l: A: NR, B: NR LDL, mmol/l: A: NR, B: NR HDL, mmol/l: A: NR, B: NR	Triglycerides, mmol/l: A: NR, B: NR Haematocrit, %: A: NR, B: NR Hb, g/dl: A: NR, B: NR Diabetes, n (%): A: NR, B: NR Other comorbidities, n (%): A: NR, B: NR SBP, mmHg: A: NR, B: NR DBP, mmHg: A: NR, B: NR PSA, ng/dl: A: NR, B: NR
First author, year: Behre 2012 Interventions: A: TRT gel (n = 183) B: Placebo (n = 179)	Age, years: A: 61.9 (6.6), B: 62.1 (6.3) BMI, kg/m²: A: 28.5 (3.3), B: 28.7 (3.0) TT, nmol/l: A: 10.4 (2.6), B: 10.6 (2.6) Free testosterone, pg/ml: A: NR, B: NR HbA1c, %: A: NR, B: NR Total cholesterol, mmol/l: A: NR, B: NR LDL, mmol/l: A: NR, B: NR HDL, mmol/l: A: NR, B: NR	Triglycerides, mmol/l: A: NR, B: NR Haematocrit, %: A: NR, B: NR Hb, g/dl: A: NR, B: NR Diabetes, n (%): A: NR, B: NR Other co-morbidities, n (%): A: NR, B: NR SBP, mmHg: A: NR, B: NR DBP, mmHg: A: NR, B: NR PSA, ng/dl: A: 1.25 (range 0.1–3.7), B: 1.31 (range 0.2–3.7)
First author, year: Borst 2014 Interventions: A: TRT IM (n = 14) B: Placebo (n = 16)	Age, years: A: 69.2 (8), B: 70.8 (9.7) BMI, kg/m²: A: 29.4 (4.6), B: 30.4 (3.4) TT, nmol/l: A: 245 (73) ng/dl Free testosterone, pg/ml: A: NR, B: NR HbA1c, %: A: NR, B: NR Total cholesterol, mmol/l: A: NR, B: NR LDL, mmol/l: A: NR, B: NR HDL, mmol/l: A: NR, B: NR	Triglycerides, mmol/l: A: NR, B: NR Haematocrit, %: A: 42.6 (3.0), B: 40.3 (3.9) Hb, g/dl: A: NR, B: NR Diabetes, n (%): A: NR, B: NR Other comorbidities, n (%): A: NR, B: NR SBP, mmHg: A: NR, B: NR DBP, mmHg: A: NR, B: NR PSA, ng/dl: A: 0.78 (0.64) ng/ml, B: 0.98 (0.53) ng/ml
First author, year: Cavallini 2004 Interventions: A: TRT oral (n = 40) B: Placebo (n = 45)	Age, years: A: 64 (range 60–72), B: 63 (range 61–74) BMI, kg/m²: A: NR, B: NR TT, nmol/l: A: 9.89 (1.84), B: 10.53 (2.11) Free testosterone, pg/ml: A: 4.4 (0.8), B: 4.2 (0.6) HbA1c, %: A: NR, B: NR Total cholesterol, mmol/l: A: NR, B: NR LDL, mmol/l: A: NR, B: NR HDL, mmol/l: A: NR, B: NR	Triglycerides, mmol/l: A: NR, B: NR Haematocrit, %: A: NR, B: NR Hb, g/dl: A: NR, B: NR Diabetes, n (%): A: NR, B: NR Other comorbidities, n (%): A: NR, B: NR SBP, mmHg: A: NR, B: NR DBP, mmHg: A: NR, B: NR PSA, ng/dl: A: 2.0 (0.7) ng/ml, B: 1.9 (0.8) ng/ml
First author, year: Cherrier 2015 Interventions: A: TRT gel (n = 10) B: Placebo (n = 12)	Age, years: A: NR, B: NR BMI, kg/m²: A: NR, B: NR TT, nmol/l: A: 308.2 (92.1) ng/dl, B: 284.2 (76.5) ng/dl Free testosterone, pg/ml: A: 4.8 (1.2) ng/dl, B: 5.0 (1.4) ng/dl HbA1c, %: A: NR, B: NR Total cholesterol, mmol/l: A: NR, B: NR LDL, mmol/l: A: NR, B: NR HDL, mmol/l: A: NR, B: NR	Triglycerides, mmol/l: A: NR, B: NR Haematocrit, %: A: NR, B: NR Hb, g/dl: A: NR, B: NR Diabetes, n (%): A: NR, B: NR Other comorbidities, n (%): A: NR, B: NR SBP, mmHg: A: NR, B: NR DBP, mmHg: A: NR, B: NR PSA, ng/dl: A: NR, B: NR
First author, year: Chiang 2007 Interventions: A: TRT gel (n = 20) B: Placebo (n = 20)	Age, years: A: 47.9 (17), B: 56.1 (14.6) BMI, kg/m²: A: NR, B: NR TT, nmol/l: A: 213.1 (158.3) ng/dl, B: 263.4 (198.1) ng/dl Free testosterone, pg/ml: A: 9.32 (17.39), B: 6.11 (4.15) HbA1c, %: A: NR, B: NR Total cholesterol, mmol/l: A: NR, B: NR LDL, mmol/l: A: NR, B: NR HDL, mmol/l: A: NR, B: NR	Triglycerides, mmol/l: A: NR, B: NR Haematocrit, %: A: NR, B: NR Hb, g/dl: A: NR, B: NR Diabetes, n (%): NR, B: NR Other comorbidities, n (%): A: NR, B: NR SBP, mmHg: A: NR, B: NR DBP, mmHg: A: NR, B: NR PSA, ng/dl: A: NR, B: NR

TABLE 36 Participant characteristics of the 18 studies that did not provide IPD (continued)

Study ID, interventions (n randomised)	Participant characteristics [all mean (SD) unless otherwise specified]	
First author, year: Clague 1999 Interventions: A: TRT IM (n = 7) B: Placebo (n = 7)	Age, years: A: 68.1 (6.6), B: 65.3 (1.8) BMI, kg/m²: A: NR, B: NR TT, nmol/l: A: 11.3 (1.7), B: 11.6 (0.9) Free testosterone, pg/ml: A: NR, B: NR HbA1c, %: A: NR, B: NR Total cholesterol, mmol/l: A: 5.66 (0.88), B: 5.57 (0.77) LDL, mmol/l: A: NR, B: NR HDL, mmol/l: A: NR, B: NR	Triglycerides, mmol/l: A: NR, B: NR Haematocrit, %: A: NR, B: NR Hb, g/dl: A: 144 (11) g/l, B: 143 (10) g/l Diabetes, n (%): A: NR, B: NR Other comorbidities, n (%): A: NR, B: NR SBP, mmHg: A: NR, B: NR DBP, mmHg: A: NR, B: NR PSA, ng/dl: A: 2.4 (1) IU/l, B: 1.9 (1.1) IU/l
First author, year: Dhindsa 2016a Interventions: A: TRT IM (n = 22) B: Placebo (n = 22)	Age, years: A: 54.7 (7.8), B: 54.5 (8.7) BMI, kg/m²: A: 39.0 (7.6), B: 39.4 (7.9) TT, nmol/l: A: 252 (82) ng/dl, B: 239 (81) ng/dl Free testosterone, pg/ml: A: 4.5 (1.3) ng/dl, B: 4.2 (1.2) ng/dl HbA1c, %: A: 6.8 (0.9), B: 7.0 (1.4) Total cholesterol, mmol/l: A: 157 (38) mg/dl, B: 156 (37) mg/dl LDL, mmol/l: A: 87 (37) mg/dl, B: 83 (23) mg/dl HDL, mmol/l: A: 34 (7) mg/dl, B: 39 (10) mg/dl	Triglycerides, mmol/l: A: 222 (197) mg/dl, B: 167 (96) mg/dl Haematocrit, %: A: NR, B: NR Hb, g/dl: A: NR, B: NR Diabetes, n (%): A: 22 (100), B: 22 (100) Other comorbidities, n (%): A: NR, B: NR SBP, mmHg: A: NR, B: NR DBP, mmHg: A: NR, B: NR PSA, ng/dl: A: NR, B: NR
First author, year: Dias 2016 Interventions: A: TRT gel (n = 16) B: Placebo (n = 13)	Age, years: A: 72 (SEM 1), B: 72 (SEM 1) BMI, kg/m²: A: 30.12 (SEM 1.11), B: 27.62 (SEM 1.15) TT, nmol/l: A: 300.05 (13.44) ng/dl, B: 303.78 (16.56) ng/dl Free testosterone, pg/ml: A: NR, B: NR HbA1c, %: A: NR, B: NR Total cholesterol, mmol/l: A: NR, B: NR LDL, mmol/l: A: NR, B: NR HDL, mmol/l: A: NR, B: NR	Triglycerides, mmol/l: A: NR, B: NR Haematocrit, %: A: NR, B: NR Hb, g/dl: A: NR, B: NR Diabetes, n (%): A: NR, B: NR Other comorbidities, n (%): A: NR, B: NR SBP, mmHg: A: 137.85 (2.61), B: 133.67 (3.72) DBP, mmHg: A: 73.30 (1.48), B: 73.11 (3.45) PSA, ng/dl: A: NR, B: NR
First author, year: Jones 2011 Interventions: A: TRT gel (n = 108) B: Placebo (n = 112)	Age, years: A: 59.9 (9.1) B: 59.9 (9.4) BMI, kg/m²: A: NR, B: NR TT, nmol/l: A: 9.2 (2.6), B: 9.5 (3.3) Free testosterone, pg/ml: A: 198.0 (49.3) pmol, B: 202.4 (62.1) pmol HbA1c, %: A: NR, B: NR Total cholesterol, mmol/l: A: NR, B: NR LDL, mmol/l: A: NR, B: NR HDL, mmol/l: A: NR, B: NR	Triglycerides, mmol/l: A: NR, B: NR Haematocrit, %: A: 0.43 (0.04) l/l, B: 0.43 (0.04) l/l Hb, g/dl: A: 14.9 (1.5), B: 14.9 (1.3) Diabetes, n (%): A: 68 (63.0), B: 69 (61.6) Other comorbidities, n (%): Metabolic syndrome: A: 88 (81.5), B: 88 (78.6) SBP, mmHg: A: NR, B: NR DBP, mmHg: A: NR, B: NR PSA, ng/dl: A: 1.6 (1.8), B: 1.2 (1.2)
First author, year: Kaufman 2011a Interventions: A: TRT gel (n = 234) B: Placebo (n = 40)	Age, years: A: 53.6 (9.5), B: 55.5 (10.3) BMI, kg/m²: A: 31.3 (4.2), B: 30.6 (4.1) TT, nmol/l: A: 282 ng/dl (SD NR), B: 294 ng/dl (SD NR) Free testosterone, pg/ml: A: NR, B: NR HbA1c, %: A: NR, B: NR Total cholesterol, mmol/l: A: NR, B: NR LDL, mmol/l: A: NR, B: NR HDL, mmol/l: A: NR, B: NR	Triglycerides, mmol/l: A: NR, B: NR Haematocrit, %: A: NR, B: NR Hb, g/dl: A: NR, B: NR Diabetes, n (%): A: NR, B: NR Other comorbidities, n (%): A: NR, B: NR SBP, mmHg: A: 129.8 (14.1), B: 130.1 (13.6) DBP, mmHg: A: NR, B: NR PSA, ng/dl: A: 0.89 (0.64), B: 0.85 (0.61)

continued

TABLE 36 Participant characteristics of the 18 studies that did not provide IPD (continued)

Study ID, interventions (n randomised)	Participant characteristics [all mean (SD) unless otherwise specified]	
First author, year: Kenny 2010 Interventions: A: TRT gel (n = 69) B: Placebo (n = 62)	Age, years: A: 77.9 (7.3), B: 76.3 (8.0) BMI, kg/m²: A: 27.2 (4.3), B: 26.6 (4.2) TT, nmol/l: A: 380.4 (179.5) ng/dl, B: 417.8 (192.5) ng/dl Free testosterone, pg/ml: A: NR, B: NR HbA1c, %: A: NR, B: NR Total cholesterol, mmol/l: A: 184 (39) mg/dl, B: 192 (37) mg/dl LDL, mmol/l: A: 117.1 (34.2) mg/dl, B: 121.0 (32.6) mg/dl HDL, mmol/l: 45.1 (11.7) mg/dl, B: 45.7 (13.4) mg/dl	Triglycerides, mmol/l: A: 118 (85) mg/dl, B: 119 (62) mg/dl Haematocrit, %: A: NR, B: NR Hb, g/dl: A: NR, B: NR Diabetes, n (%): A: 12 (17.4), B: 10 (16.1) Other comorbidities, n (%): A: Hypertension: 22 (31.9%), coronary artery disease: 24 (34.8%), allergies: 16 (23.2%), cancer: 15 (21.7%), depression: 15 (21.7%), muscle aches: 21 (30.4%), hearing problems: 36 (52.2%), osteoarthritis: 19 (27.5%), joint pain: 33 (47.8%) B: Hypertension: 24 (38.7%), coronary artery disease: 36 (58.1%), allergies: 21 (33.9%), cancer: 5 (8.1%), depression: 11 (17.7%), muscle aches: 26 (41.9%), hearing problems: 21 (33.9%), osteoarthritis: 21 (33.9%), joint pain: 29 (46.8%) SBP, mmHg: A: NR, B: NR DBP, mmHg: A: NR, B: NR PSA, ng/dl: A: 2.1 (1.5), B: 2.0 (1.3)
First author, year: Morales 2009 Interventions: A: TRT capsules (n = 29) B: Placebo (n = 29)	Age, years: A: 59 (10.6), B: 60.2 (9.6) BMI, kg/m²: A: 31.3 (5.4), B: 29.7 (4.4) TT, nmol/l: A: 10.2 (4.9), B: 10.0 (5.5) Free testosterone, pg/ml: A: NR, B: NR HbA1c, %: A: NR, B: NR Total cholesterol, mmol/l: A: NR, B: NR LDL, mmol/l: A: NR, B: NR HDL, mmol/l: A: NR, B: NR	Triglycerides, mmol/l: A: NR, B: NR Haematocrit, %: A: NR, B: NR Hb, g/dl: A: NR, B: NR Diabetes, n (%): A: NR, B: NR Other comorbidities, n (%): A: NR, B: NR SBP, mmHg: A: NR, B: NR DBP, mmHg: A: NR, B: NR PSA, ng/dl: A: NR, B: NR
First author, year: Paduch 2015a Interventions: A: TRT solution (n = 40) B: Placebo (n = 36)	Age, years: A: 48.4 (9.8), B: 52.7 (9.3) BMI, kg/m²: A: 30.6 (3.05), B: 30.8 (3.16) TT, nmol/l: A: 214 (56) ng/dl, B: 223 (53) ng/dl Free testosterone, pg/ml: A: 5.3 (2.0) ng/dl, B: 5.6 (1.6) ng/dl HbA1c, %: A: 5.9 (0.6), B: 6.1 (1.2) Total cholesterol, mmol/l: A: 191 (43) mg/dl, B: 204 (43) mg/dl LDL, mmol/l: A: 111 (39) mg/dl, B: 120 (41) mg/dl HDL, mmol/l: A: 50 (11) mg/dl, B: 49 (12) mg/dl	Triglycerides, mmol/l: A: 150 (44) mg/dl, B: 190 (168) mg/dl Haematocrit, %: A: 42.6 (4.0), B: 42.6 (3.1) Hb, g/dl: A: 14.4 (1.3), B: 14.7 (1.1) Diabetes, n (%): A: NR, B: NR Other comorbidities, n (%): A: Reduced ejaculate volume: 31 (86.1); delayed ejaculation: 24 (66.7); anejaculation: 14 (38.9); decreased ejaculate force: 29 (80.6) B: Reduced ejaculate volume: 31 (77.5); delayed ejaculation: 20 (50); anejaculation: 11 (27.5); decreased ejaculate force: 31 (77.5) SBP, mmHg: A: NR, B: NR DBP, mmHg: A: NR, B: NR PSA, ng/dl: A: 0.92 (0.75) ng/ml, B: 0.86 (0.55) ng/ml
First author, year: Steidle 2003 Interventions: A: TRT gel 50 mg (n = 99) B: TRT gel 100 mg (n = 106) C: TRT patch (n = 102) D: Placebo (n = 99)	Age, years: A: 58.1 (9.7), B: 56.8 (10.6), C: 60.5 (9.7), D: 56.8 (10.8) BMI, kg/m²: A: 30.0 (3.7), B: 29.9 (3.3), C: 29.9 (3.8), D: 30.3 (3.8) TT, nmol/l: A: 8.1 (2.0), B: 8.1 (2.2), C: 8.3 (2.4), D: 7.9 (2.8) Free testosterone, pg/ml: A: NR, B: NR, C: NR, D: NR HbA1c, %: A: NR, B: NR, C: NR, D: NR Total cholesterol, mmol/l: A: NR, B: NR, C: NR, D: NR LDL, mmol/l: A: NR, B: NR, C: NR, D: NR HDL, mmol/l: A: NR, B: NR, C: NR, D: NR	Triglycerides, mmol/l: A: NR, B: NR, C: NR, D: NR Haematocrit, %: A: NR, B: NR, C: NR, D: NR Hb, g/dl: A: NR, B: NR, C: NR, D: NR Diabetes, n (%): A: NR, B: NR, C: NR, D: NR Other comorbidities, n (%): A: NR, B: NR, C: NR, D: NR SBP, mmHg: A: NR, B: NR, C: NR, D: NR DBP, mmHg: A: NR, B: NR, C: NR, D: NR PSA, ng/dl: A: 1.17 (0.89), B: 1.29 (0.96), C: 1.45 (1.18), D: 1.13 (1.00)

TABLE 36 Participant characteristics of the 18 studies that did not provide IPD (continued)

Study ID, interventions (n randomised)	Participant characteristics [all mean (SD) unless otherwise specified]	
First author, year: Wang 2013 Interventions: A: TRT capsules 40 mg (n = 62) B: TRT capsules 20 mg (n = 62) C: Placebo (n = 62)	Age, years: A: 68.1 (5.4), B: 68.4 (5.5), C: 68.0 (4.8) BMI, kg/m²: A: 27.9 (3.2), B: 28.2 (3.6), C: 28.7 (2.9) TT, nmol/l: A: 214.8 (22.4) ng/dl, B: 218.3 (25.1) ng/dl, C: 220.1 (20.7) ng/dl Free testosterone, pg/ml: A: 4.2 (1.1), B: 3.9 (0.9), C: 3.8 (0.7) HbA1c, %: A: 6.6 (0.7), B: 6.8 (0.6), C: 6.4 (0.8) Total cholesterol, mmol/l: A: 4.9 (1.2), B: 4.4 (1.0), C: 5.1 (1.6) LDL, mmol/l: A: 3.3 (0.9), B: 3.1 (0.6), C: 3.4 (1.2) HDL, mmol/l: A: 0.9 (0.2), B: 0.9 (0.3), C: 0.8 (0.2)	Triglycerides, mmol/l: A: 2.9 (1.2), B: 2.6 (1.3), C: 3.1 (1.4) Haematocrit, %: A: NR, B: NR, C: NR Hb, g/dl: A: NR, B: NR, C: NR Diabetes, n (%): A: NR, B: NR, C: NR Other comorbidities, n (%): Osteoporosis: A: 62 (100), B: 62 (100), C: 62 (100) SBP, mmHg: A: 136.2 (15.8), B: 138.5 (9.9), C: 142.8 (12.8) DBP, mmHg: A: 82.1 (4.5), B: 86.2 (5.6), C: 87.1 (6.2) PSA, ng/dl: A: 3.9 (0.8) ng/ml, B: 3.6 (1.0) ng/ml, C: 3.7 (0.7) ng/ml
IQR, inter-quartile range; IM, intramuscular; MS, metabolic syndrome; NR, not reported.		

Appendix 6 Clinical effectiveness results

TABLE 37 Baseline characteristics of patients in studies that contributed to the IPD

Baseline characteristic	Number of studies	TRT mean (SD), n	Placebo mean (SD), n
WHOLQOL-OLD	1		
Total		71.93 (7.20); 19	70.96 (9.04); 17
Sensory ability		78.27 (16.13); 21	76.39 (15.24); 18
Autonomy		68.44 (14.55); 20	65.79 (20.66); 19
Past, present and future activities		68.75 (9.93); 20	69.10 (11.83); 18
Social participation		66.88 (14.78); 20	67.71 (15.93); 18
Death and dying		69.69 (17.36); 20	72.57 (12.71); 18
Intimacy		76.19 (12.12); 21	73.96 (12.73); 18
Herschbach questionnaire	1		
General		73.91 (25.44); 110	79.77 (25.24); 110
Health		80.87 (28.16); 110	78.27 (34.20); 110
Hormones		107.59 (56.80); 107	113.55 (63.10); 107
Eleven questions about sexual functioning	1	0.82 (0.39); 111	0.77 (0.42); 110
PDQ	1		
Activity		1.47 (1.33); 387	1.53 (1.40); 384
Positive mood		4.86 (1.08); 391	4.83 (1.06); 391
Negative mood		1.45 (0.94); 391	1.41 (0.90); 391
Sexual desire		1.96 (1.40); 391	1.97 (1.31); 391
% of full erection		42.46 (21.60); 229	44.98 (21.34); 221
Sexual enjoyment with partner		0.85 (0.96); 391	0.82 (0.94); 391
Sexual enjoyment without partner		1.41 (1.54); 308	1.49 (1.66); 315
Satisfaction		2.42 (1.92); 229	2.68 (1.91); 221
DISF-II5	1		
Total		39.19 (22.56); 251	38.65 (24.29); 258
Sexual cognition and fantasy		12.10 (6.84); 252	11.99 (6.96); 260
Sexual arousal		5.16 (5.51); 253	5.20 (5.64); 259
Sexual behaviour and experience		6.67 (4.79); 252	6.86 (4.89); 259
Orgasm		5.34 (5.33); 253	5.72 (5.74); 260
Sexual drive and relationship		9.90 (6.89); 252	9.04 (6.78); 259
Areal bone mineral density (g/cm ²)			
Total	4	1.21 (0.14); 416	1.21 (0.12); 380
Subtotal	2	1.03 (0.10); 126	1.02 (0.09); 120
Femoral neck	6	0.84 (0.14); 270	0.86 (0.16); 231
Lumbar spine	10	1.17 (0.21); 550	1.17 (0.23); 512

continued

TABLE 37 Baseline characteristics of patients in studies that contributed to the IPD (*continued*)

Baseline characteristic	Number of studies	TRT mean (SD), n	Placebo mean (SD), n
Thoracic spine	2	0.95 (0.15); 180	0.91 (0.14); 157
Total hip	9	1.02 (0.14); 420	1.03 (0.15); 380
Trochanter	3	0.78 (0.12); 97	0.76 (0.11); 70
Intertrochanter	2	1.15 (0.16); 37	1.16 (0.16); 37
Pelvis	2	1.23 (0.19); 175	1.20 (0.16); 140
Left arm	3	0.85 (0.11); 202	0.83 (0.07); 176
Right arm	3	0.86 (0.12); 202	0.84 (0.08); 177
Left plus right arm	1	1.63 (0.12); 20	1.57 (0.40); 19
Left leg	3	1.26 (0.17); 192	1.23 (0.11); 159
Right leg	3	1.26 (0.15); 188	1.26 (0.18); 154
Left plus right leg	1	2.42 (0.22); 20	2.30 (0.63); 18
Left rib	2	0.71 (0.08); 179	0.69 (0.07); 155
Right rib	2	0.72 (0.08); 182	0.69 (0.07); 159
Head	2	2.11 (0.30); 183	2.06 (0.33); 159
Shaft	1	1.21 (0.17); 54	1.17 (0.15); 28
Wards	1	0.60 (0.16); 54	0.54 (0.12); 28
Volumetric bone mineral density (mg/cm ³)	1		
Spine trabecular		102.39 (31.91); 110	99.37 (26.95); 97
Spine cortical		285.40 (42.47); 110	284.18 (43.31); 97
Spine whole		193.35 (37.24); 110	192.64 (34.90); 97
Hips trabecular		185.42 (34.32); 103	180.72 (33.05); 88
Hips cortical		398.99 (46.36); 103	391.68 (50.22); 88
Hips whole		248.78 (37.86); 103	243.06 (38.86); 88
Positive and Negative Affect Scale	1		
Total		23.25 (4.00); 382	23.19 (4.16); 389
Positive		16.16 (3.58); 382	16.03 (3.56); 389
Negative		7.09 (2.46); 382	7.16 (2.69); 389
Hospital Anxiety and Depression Scale (Depression)	1	7.97 (3.80); 86	7.24 (4.09); 103
Patient Health Questionnaire-9	1	5.39 (3.93); 383	5.35 (3.99); 388
Centre for Epidemiologic Studies Depression Scale	1	5.76 (5.53); 55	5.61 (5.01); 28
Aggression questionnaire	1		
Total		59.05 (11.62); 55	58.54 (13.65); 28
Physical aggression		17.13 (5.69); 55	17.32 (4.67); 28
Verbal aggression		13.31 (3.40); 55	12.36 (3.38); 28
Anger		14.82 (4.16); 55	14.11 (4.36); 28
Hostility		13.80 (3.58); 55	14.11 (4.36); 28
Spielberger State-Trait Anxiety	1	24.67 (4.98); 24	32.52 (9.05); 23

Note

Values are mean (standard deviation); numbers or numbers (per cent).

TABLE 38 Two-stage IPD meta-analysis for mortality from any cause

Outcome	Number of studies	TRT n/N (%) or n (%)	Placebo n/N (%) or n (%)
Mortality from any cause ^a	21	11/2367 (0.5)	18/2091 (0.9)
Details		N = 11	N = 18
Myocardial infarction	4	2 (18.2)	3 (16.7)
Cancer	1	0 (0)	3 (16.7)
Ruptured aortic aneurysm	1	0 (0)	1 (5.5)
Constrictive pericarditis	1	1 (9.1)	0 (0)
Coronary heart disease	1	1 (9.1)	0 (0)
Multiple organ failure	1	1 (9.1)	0 (0)
Arrhythmia	1	1 (9.1)	0 (0)
Postoperative septicaemia	1	0 (0)	1 (5.5)
Venous thromboembolism	1	0 (0)	1 (5.5)
Unknown	3	5 (45.5)	9 (50)

a Of the 21 studies, 11 reported no deaths (Amory 2004, Brock 2016, Emmelot-Vonk 2008, Giantti 2014a, Groti 2018, Hildreth 2013, Kaufman 2011a, Magnussen 2016, Merza 2006, Paduch 2015a and Srinivas-Shankar 2010).

Note

Values are numbers (per cent) or numbers.

TABLE 39 Two-stage IPD meta-analysis for CV and/or CBV events

Outcome	Number of studies	TRT n/N (%)	Placebo n/N (%)
CV and/or CBV events ^a	22	123/2496 (4.9)	116/2073 (5.6)
Number of participants with a CV event	20	110/123 (89.4)	111/116 (95.7)
Total number of CV events ^b	20	169	182
Details			
Arrhythmia	7	53	47
Coronary heart disease	7	34	33
Heart failure	7	23	28
Myocardial infarction	10	10	19
Valvular heart disease	2	18	12
Peripheral vascular disease	4	8	14
Stable angina	5	7	7
Aortic aneurysm	6	6	8
New angina	3	5	5
Unstable angina	3	2	4

continued

TABLE 39 Two-stage IPD meta-analysis for CV and/or CBV events (continued)

Outcome	Number of studies	TRT n/N (%)	Placebo n/N (%)
Aortic dissection	1	2	0
Atherosclerosis	1	1	1
Cardiac arrest	2	0	2
Angina	1	0	1
Open heart surgery	1	0	1
Number of participants with a CBV event	11	15/120 (12.5)	7/110 (6.4)
Total number of CBV events ^b	11	16	7

a Of the 22 studies, five (Groti 2018, Kaufman 2011a, Magnussen 2016, Paduch 2015a and Steidle 2003) reported no CV and/or CBV events.

b Some participants had more than one event.

Note

Values are numbers (per cent) or numbers.

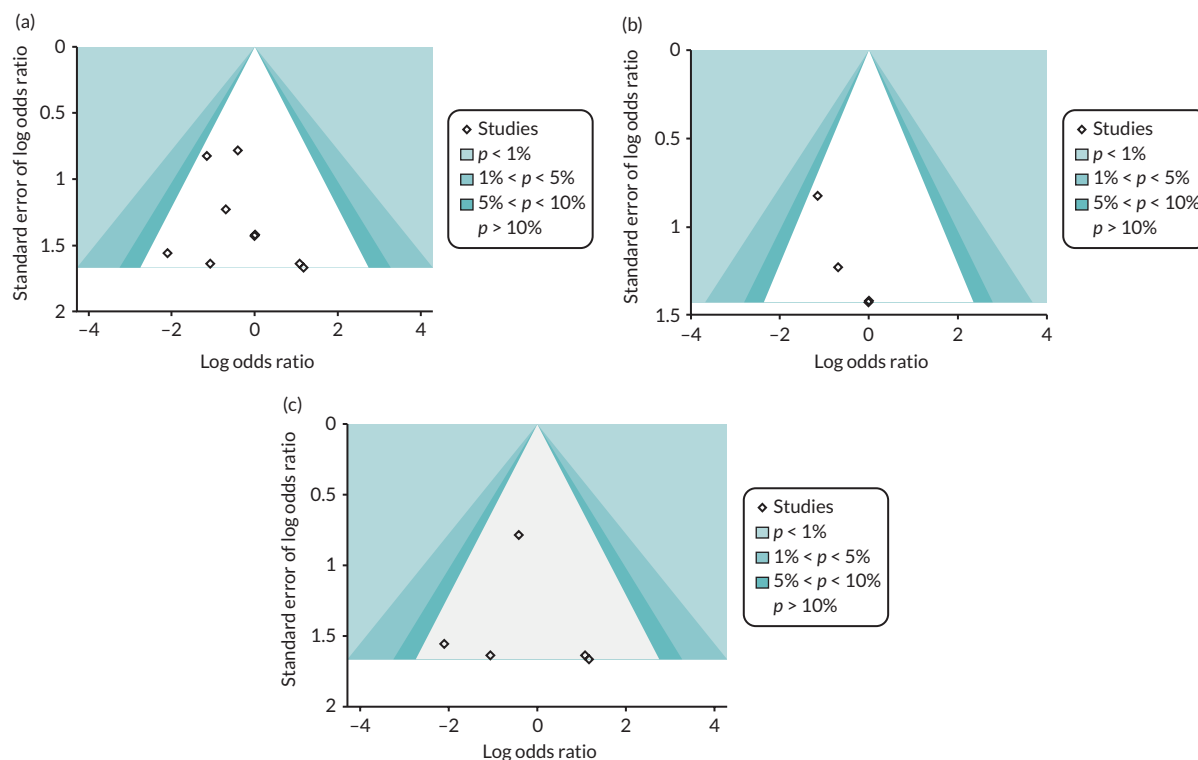


FIGURE 21 Contour-enhanced funnel plots for mortality from any cause. (a) IPD and aggregate data; (b) IPD only; and (c) aggregate data only.

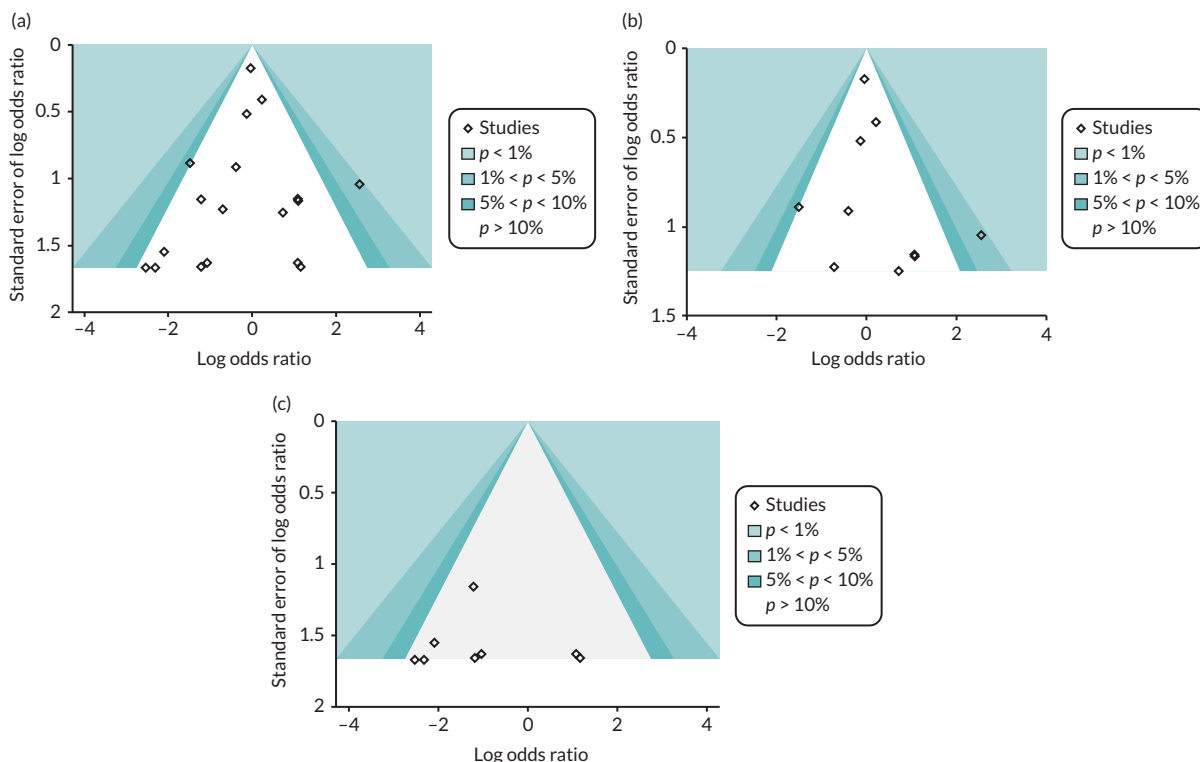


FIGURE 22 Contour-enhanced funnel plots for CV and/or CBV events. (a) IPD and aggregate data; (b) IPD only; and (c) aggregate data only.

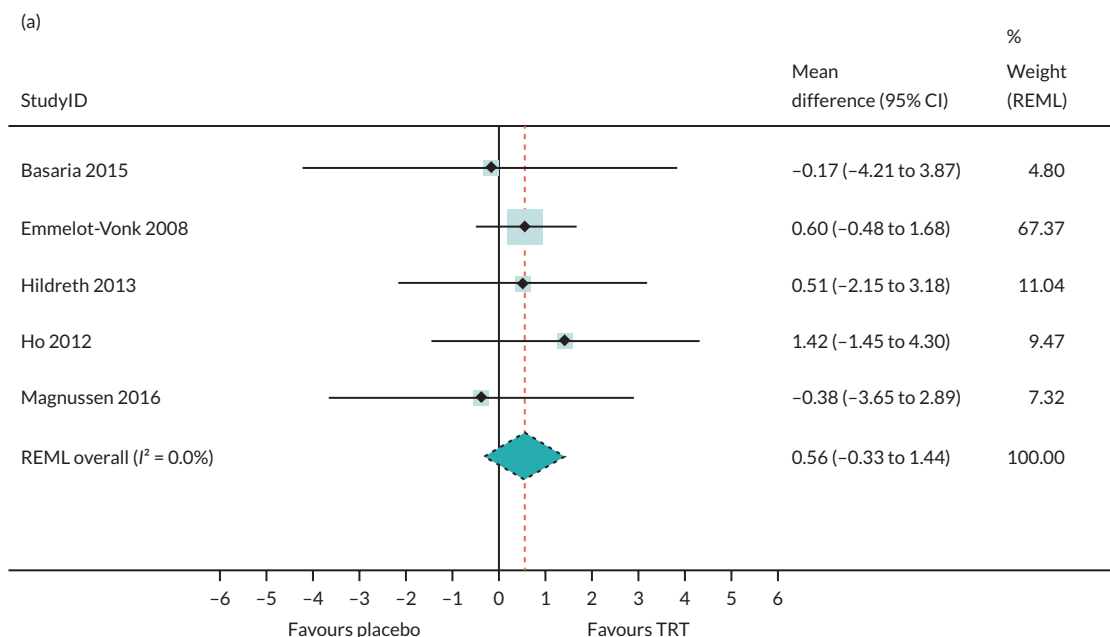


FIGURE 23 Two-stage meta-analysis for SF-36/SF-12 norm-based scores. (a) Physical Functioning; (b) Role-Physical; (c) Bodily Pain; (d) General Health; (e) Vitality; (f) Social Functioning; (g) Role Emotional; (h) Mental Health; (i) Physical Health Composite score; (j) Mental Health Composite Score.

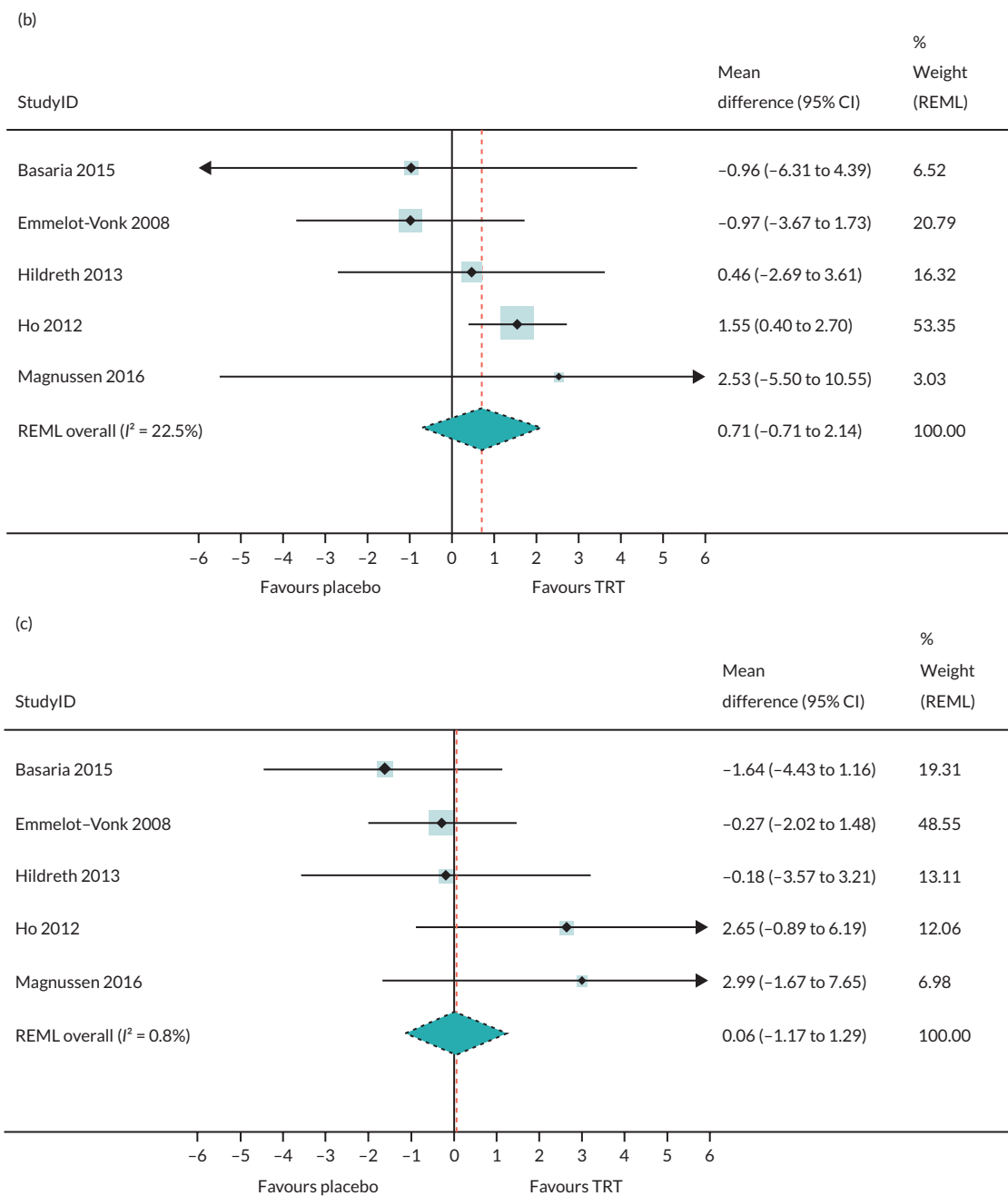


FIGURE 23 Two-stage meta-analysis for SF-36/SF-12 norm-based scores. (a) Physical Functioning; (b) Role-Physical; (c) Bodily Pain; (d) General Health; (e) Vitality; (f) Social Functioning; (g) Role Emotional; (h) Mental Health; (i) Physical Health Composite score; (j) Mental Health Composite Score. (continued)

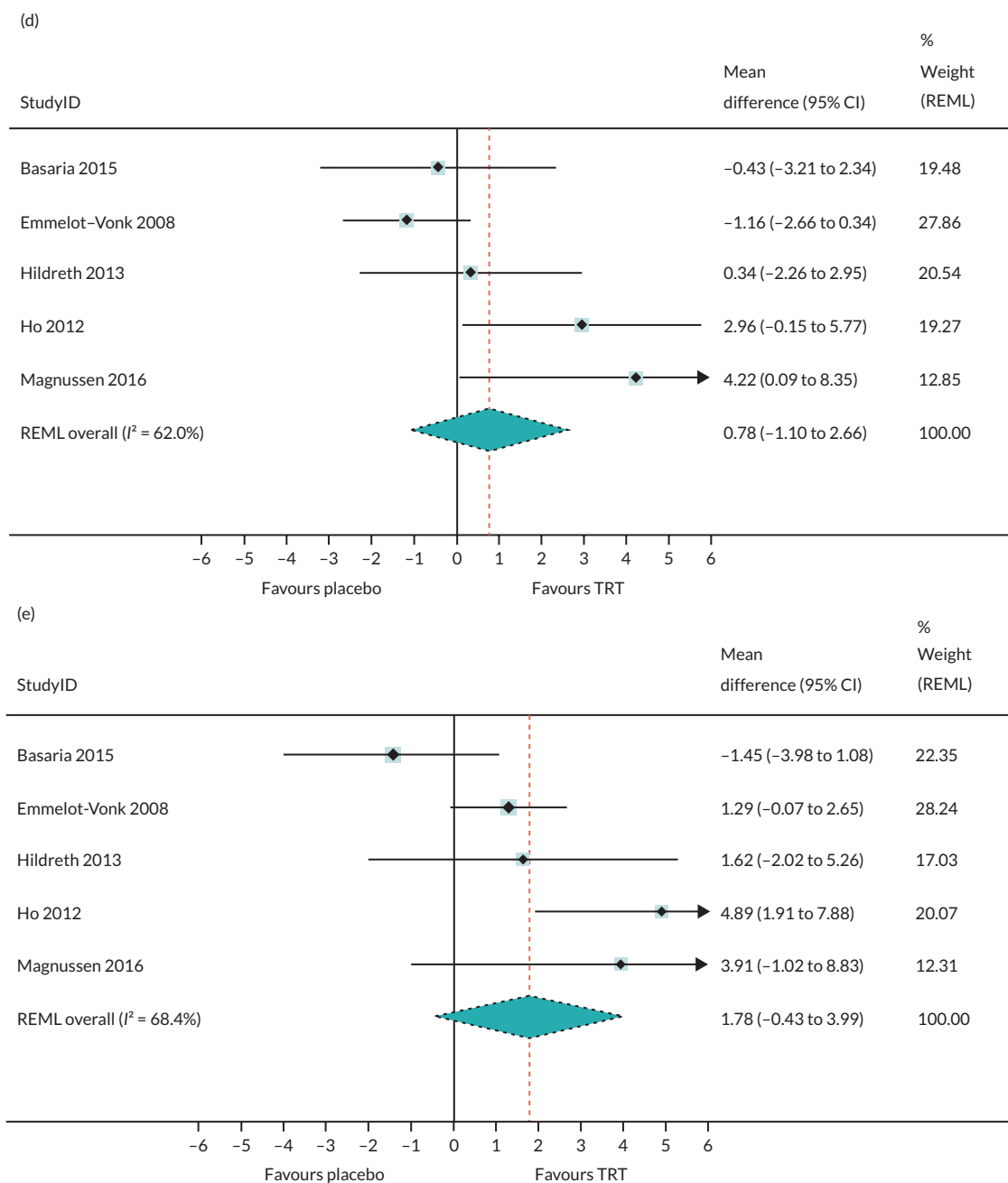


FIGURE 23 Two-stage meta-analysis for SF-36/SF-12 norm-based scores. (a) Physical Functioning; (b) Role-Physical; (c) Bodily Pain; (d) General Health; (e) Vitality; (f) Social Functioning; (g) Role Emotional; (h) Mental Health; (i) Physical Health Composite score; (j) Mental Health Composite Score. (continued)

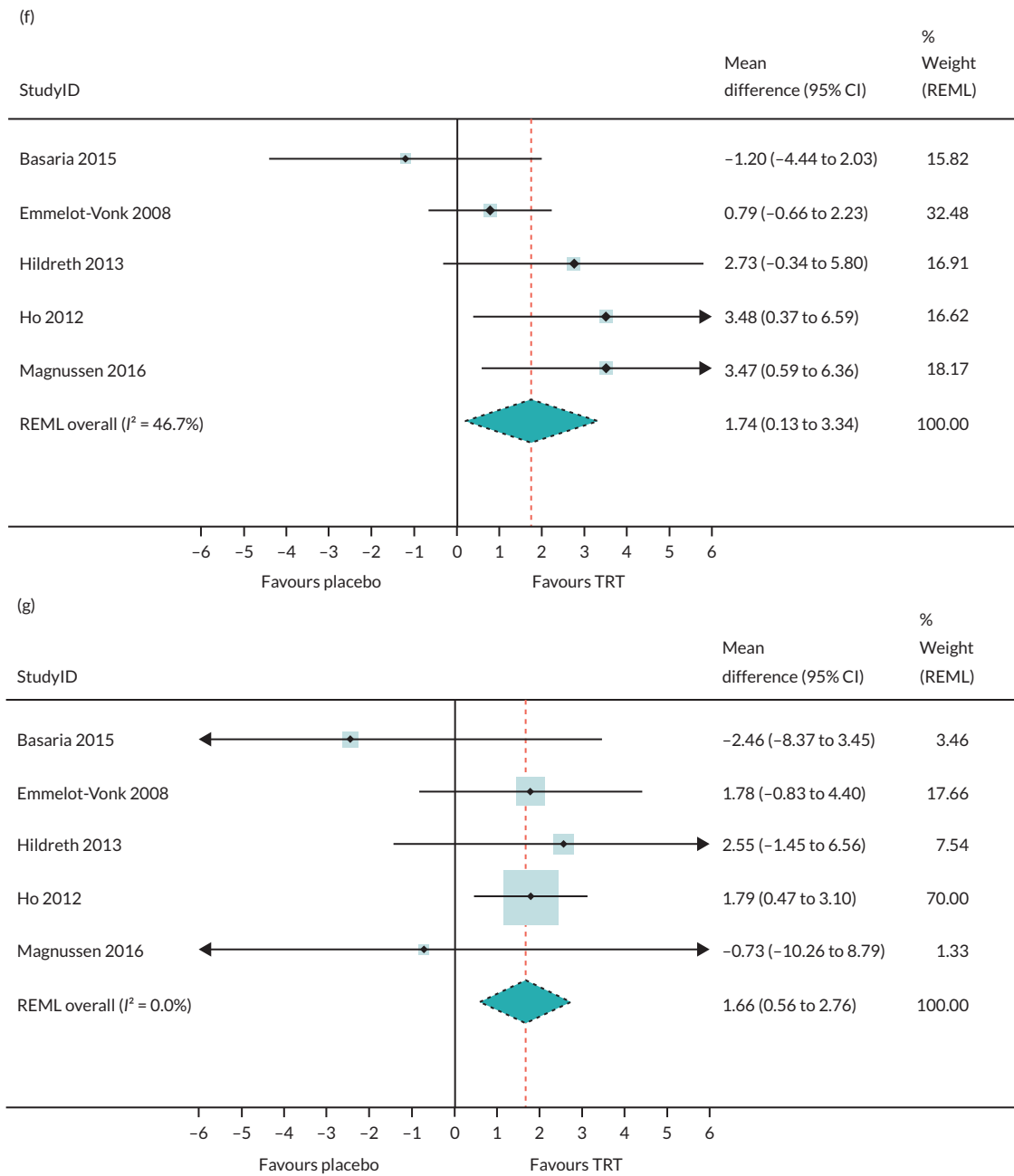


FIGURE 23 Two-stage meta-analysis for SF-36/SF-12 norm-based scores. (a) Physical Functioning; (b) Role-Physical; (c) Bodily Pain; (d) General Health; (e) Vitality; (f) Social Functioning; (g) Role Emotional; (h) Mental Health; (i) Physical Health Composite score; (j) Mental Health Composite Score. (continued)

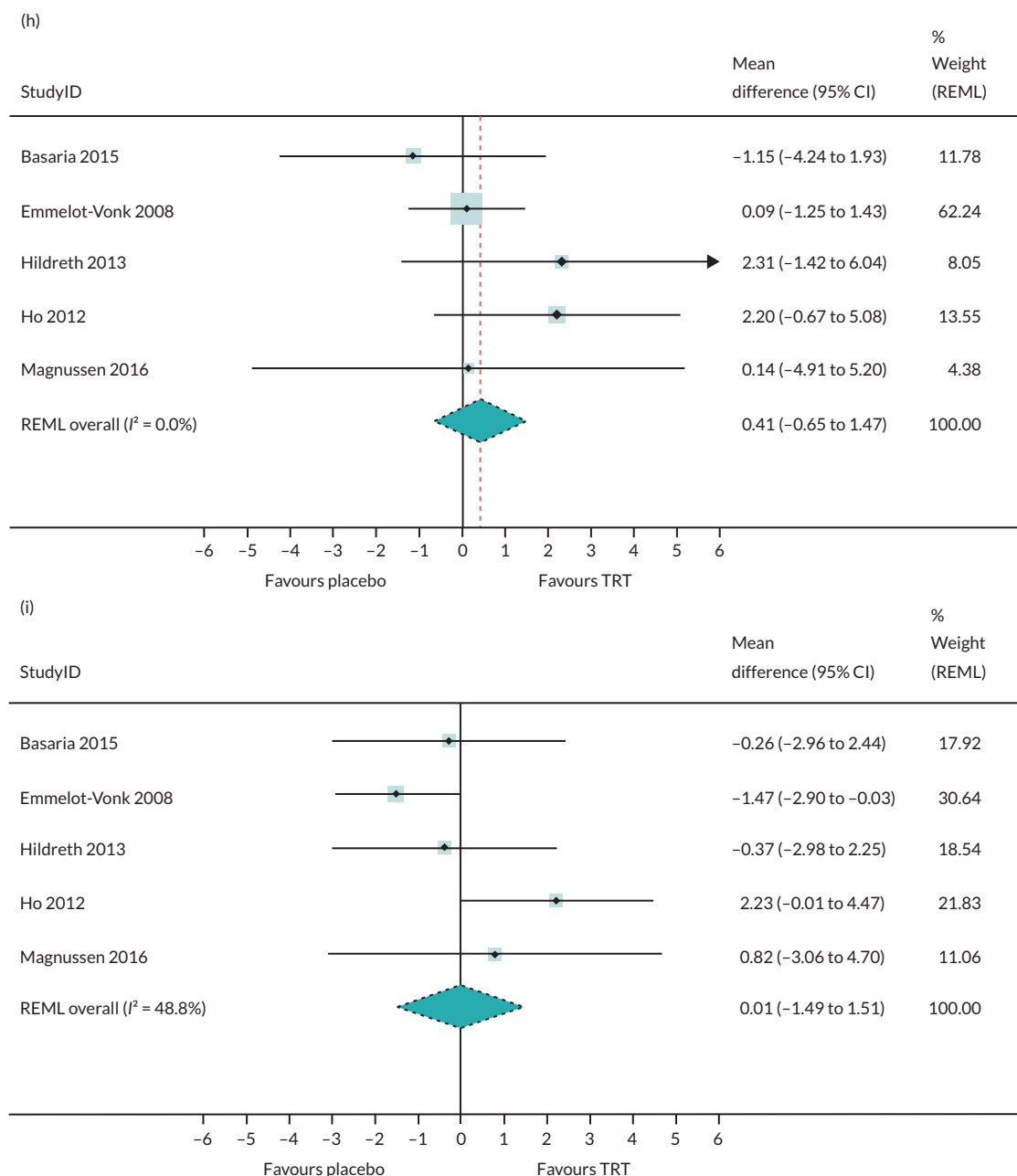


FIGURE 23 Two-stage meta-analysis for SF-36/SF-12 norm-based scores. (a) Physical Functioning; (b) Role-Physical; (c) Bodily Pain; (d) General Health; (e) Vitality; (f) Social Functioning; (g) Role Emotional; (h) Mental Health; (i) Physical Health Composite score; (j) Mental Health Composite Score. (*continued*)

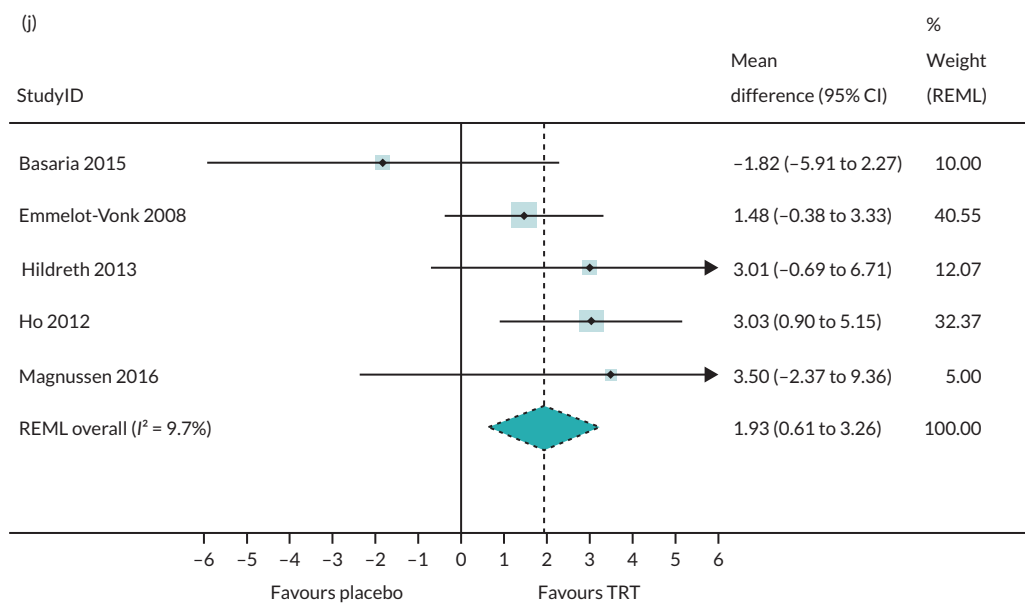


FIGURE 23 Two-stage meta-analysis for SF-36/SF-12 norm-based scores. (a) Physical Functioning; (b) Role-Physical; (c) Bodily Pain; (d) General Health; (e) Vitality; (f) Social Functioning; (g) Role Emotional; (h) Mental Health; (i) Physical Health Composite score; (j) Mental Health Composite Score.

TABLE 40 One-stage analysis for QoL outcomes

Outcome	Number of studies	TRT mean (SD), n	Placebo mean (SD), n	MD	95% CI
WHOLQOL-OLD					
Total	1	72.09 (7.29); 19	72.09 (9.67); 19	-0.26	(-3.52 to 3.00)
Sensory ability	1	78.95 (14.31); 19	76.32 (16.48); 19	0.34	(-7.79 to 8.48)
Autonomy	1	69.08 (11.69); 19	71.05 (12.01); 19	-1.96	(-7.60 to 3.68)
Past, present and future activities	1	64.80 (8.13); 19	69.41 (12.65); 19	-3.68	(-10.01 to 2.64)
Social participation	1	67.43 (12.07); 19	70.07 (14.22); 19	-0.93	(-6.57 to 4.71)
Death and dying	1	74.67 (15.23); 19	71.05 (16.17); 19	4.77	(-2.77 to 12.31)
Intimacy	1	77.63 (9.39); 19	74.67 (15.23); 19	1.11	(-4.13 to 6.34)
Herschbach questionnaire					
General	1	74.39 (26.50); 109	79.47 (26.84); 105	-0.11	(-4.91 to 4.68)
Health	1	75.92 (28.67); 112	77.69 (30.74); 105	-3.19	(-8.76 to 2.38)
Hormones	1	112.44 (56.87); 110	106.80 (58.15); 107	12.42	(3.10 to 21.75)

Note
A fixed-effects analysis was used as only one study reported the outcome.

TABLE 41 One-stage analysis for sexual function outcomes

Outcome	Number of studies	TRT mean (SD), n	Placebo mean (SD), n	MD	95% CI
Androgen deficiency in ageing males PDQ	1	3.35 (2.39); 113	3.32 (2.36); 108	-0.25	(-0.73 to 0.23)
Positive mood	1	5.26 (1.06); 210	5.06 (1.07); 208	0.18	(0.03 to 0.32)
Negative mood	1	1.12 (0.88); 210	1.21 (0.90); 208	-0.14	(-0.26 to -0.02)
Activity	1	1.64 (1.74); 199	1.41 (1.63); 201	0.30	(0.02 to 0.58)
Sexual desire	1	2.10 (1.65); 210	1.71 (1.50); 208	0.50	(0.27 to 0.73)
Sexual enjoyment with partner	1	1.01 (1.22); 210	0.71 (1.02); 208	0.35	(0.17 to 0.53)
Sexual enjoyment without partner	1	1.55 (1.79); 186	1.37 (1.71); 185	0.25	(-0.03 to 0.53)
Satisfaction	1	3.15 (1.99); 121	3.38 (2.17); 104	-0.20	(-0.73 to 0.33)
% of full erection	1	47.93 (21.03); 121	51.02 (23.23); 104	-2.35	(-8.44 to 3.74)
DISF-II5					
Total	1	50.66 (27.58); 225	40.74 (27.22); 225	9.62	(6.08 to 13.16)
Sexual cognition and fantasy	1	14.61 (6.91); 226	12.03 (7.01); 228	2.52	(1.51 to 3.53)
Sexual arousal	1	8.44 (7.52); 226	6.15 (6.77); 226	2.44	(1.56 to 3.33)
Sexual behaviour and experience	1	8.99 (5.87); 226	7.24 (5.06); 228	1.97	(1.20 to 2.75)
Orgasm	1	7.47 (6.05); 225	6.20 (6.17); 228	1.65	(0.82 to 2.49)
Sexual drive and relationship	1	11.25 (7.35); 226	9.06 (7.29); 227	1.50	(0.39 to 2.62)
Eleven questions about sexual functioning	1	0.81 (0.40); 113	0.81 (0.39); 108	-0.04	(-0.12 to 0.04)

Notes

Values are mean (standard deviation); numbers.

A fixed-effects analysis was used as only one study reported the outcome.

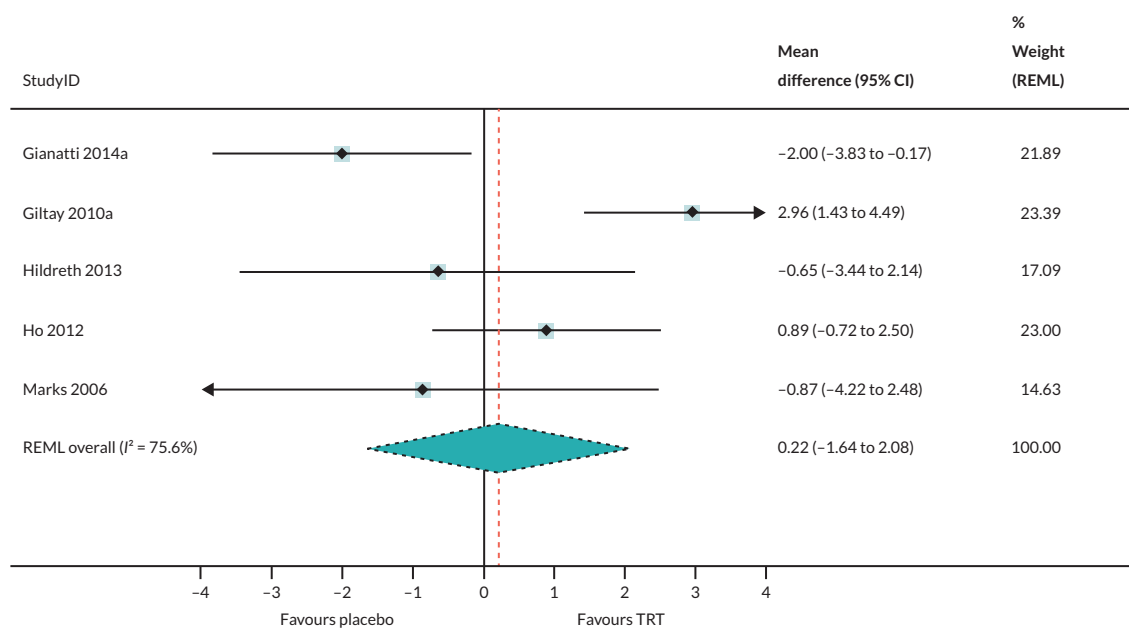


FIGURE 24 Two-stage meta-analysis for IIEF-5.

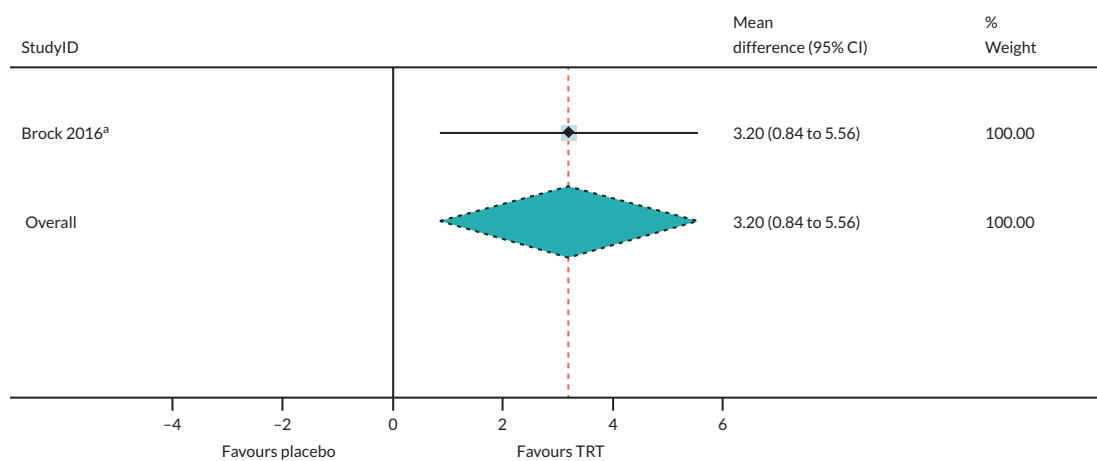


FIGURE 25 Two-stage meta-analysis for HED. a, Data presented as change from baseline.

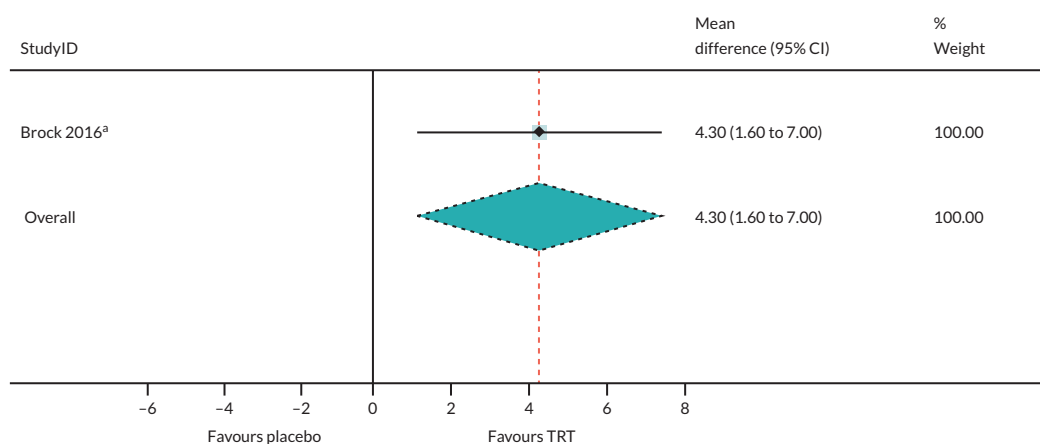


FIGURE 26 Two-stage meta-analysis for Sexual Arousal, Interest, and Drive Scale. a, Data presented as change from baseline.

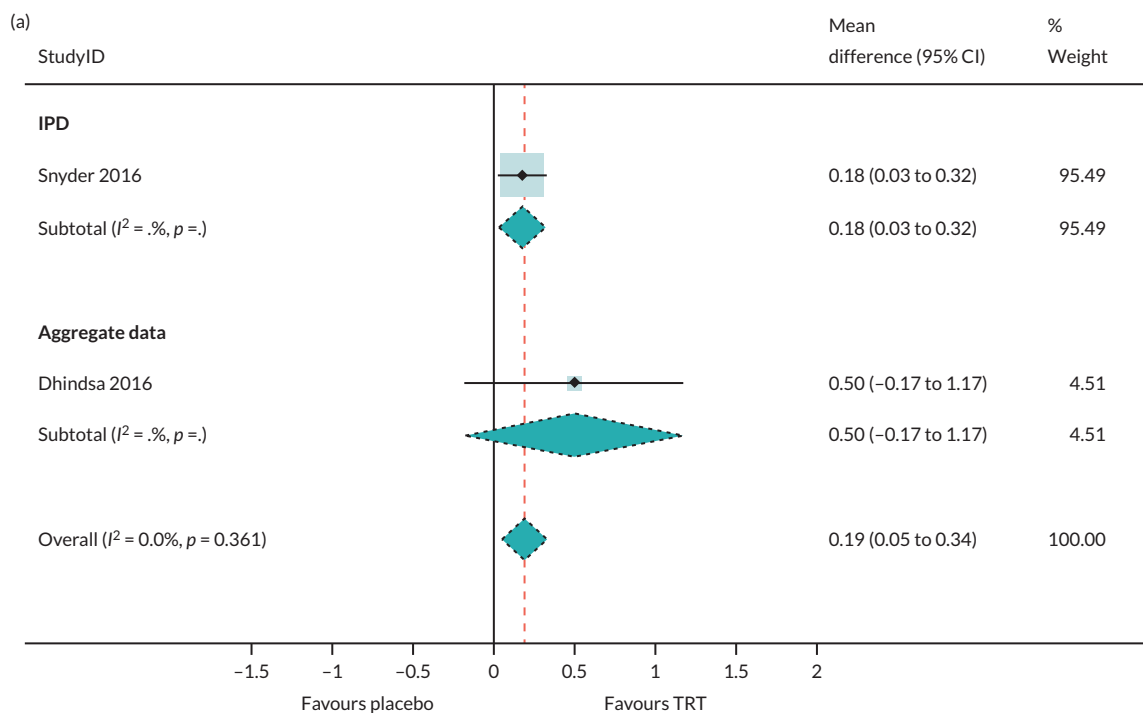


FIGURE 27 Two-stage meta-analysis for PDQ. (a) Positive mood; (b) negative mood; (c) sexual desire; (d) sexual enjoyment with partner; (e) sexual enjoyment without partner; (f) Satisfaction; (g) Percentage of full erection. a, Data presented as change from baseline. (continued)

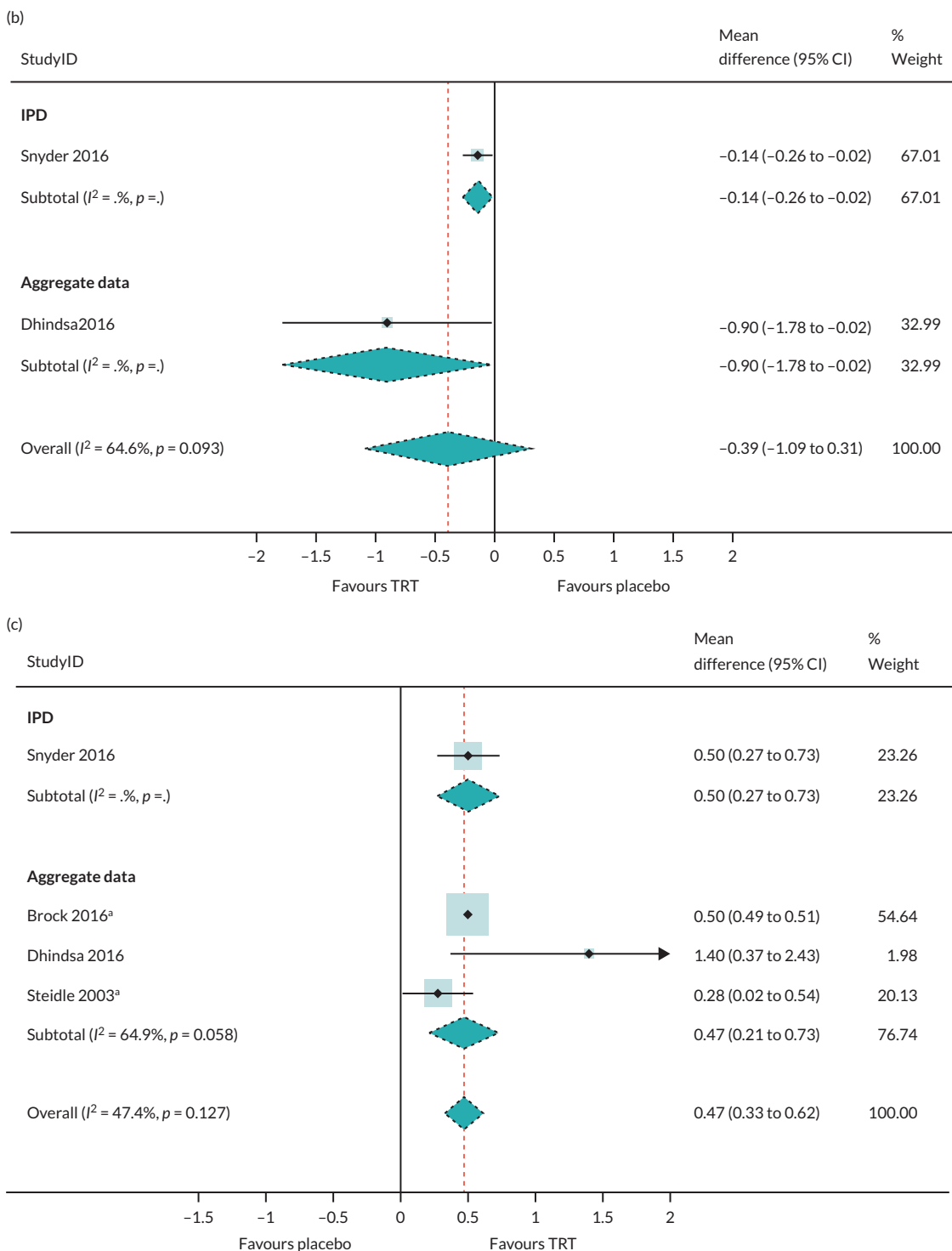


FIGURE 27 Two-stage meta-analysis for PDQ. (a) Positive mood; (b) negative mood; (c) sexual desire; (d) sexual enjoyment with partner; (e) sexual enjoyment without partner; (f) Satisfaction; (g) Percentage of full erection. a, Data presented as change from baseline. (continued)

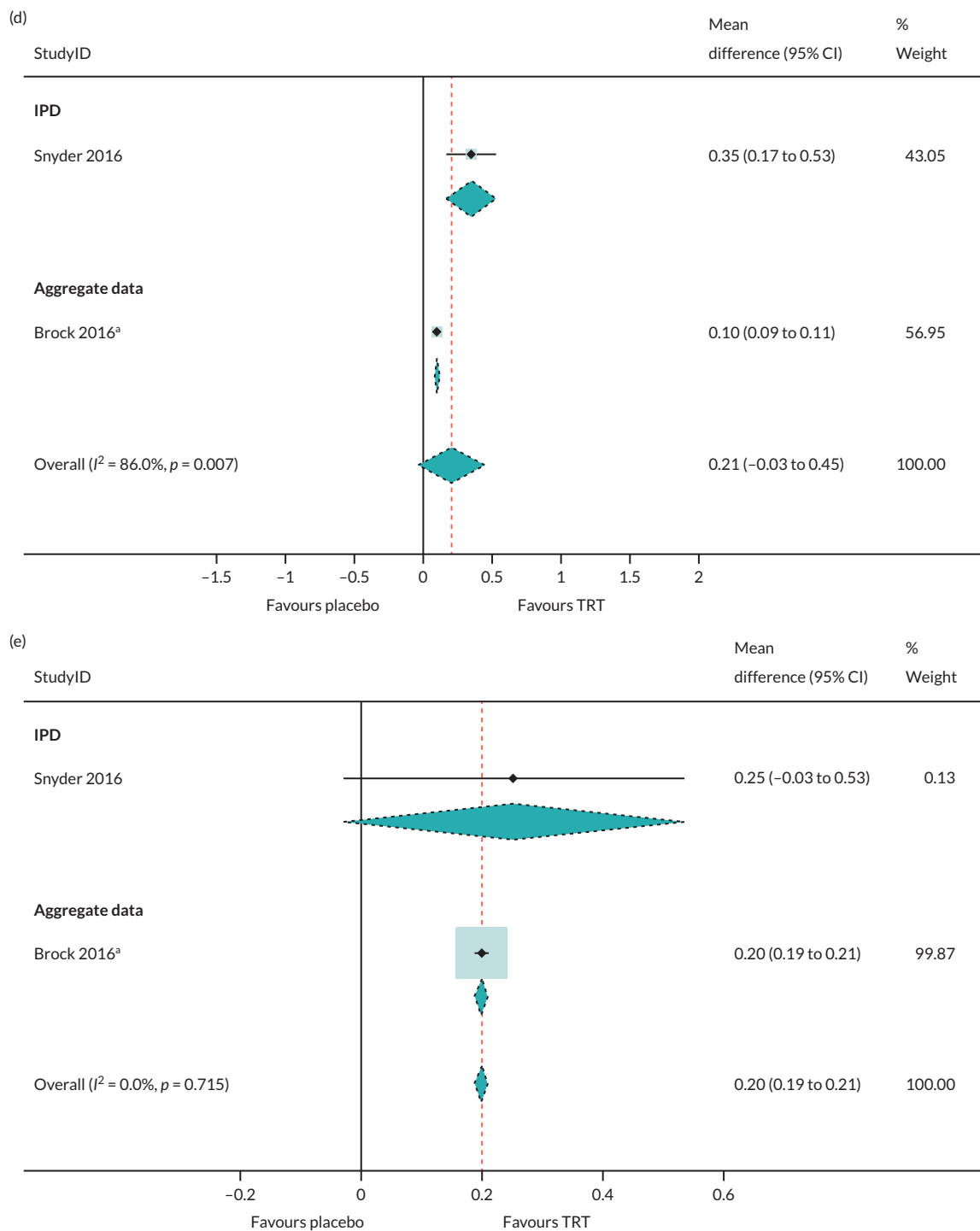


FIGURE 27 Two-stage meta-analysis for PDQ. (a) Positive mood; (b) negative mood; (c) sexual desire; (d) sexual enjoyment with partner; (e) sexual enjoyment without partner; (f) Satisfaction; (g) Percentage of full erection. a, Data presented as change from baseline. (continued)

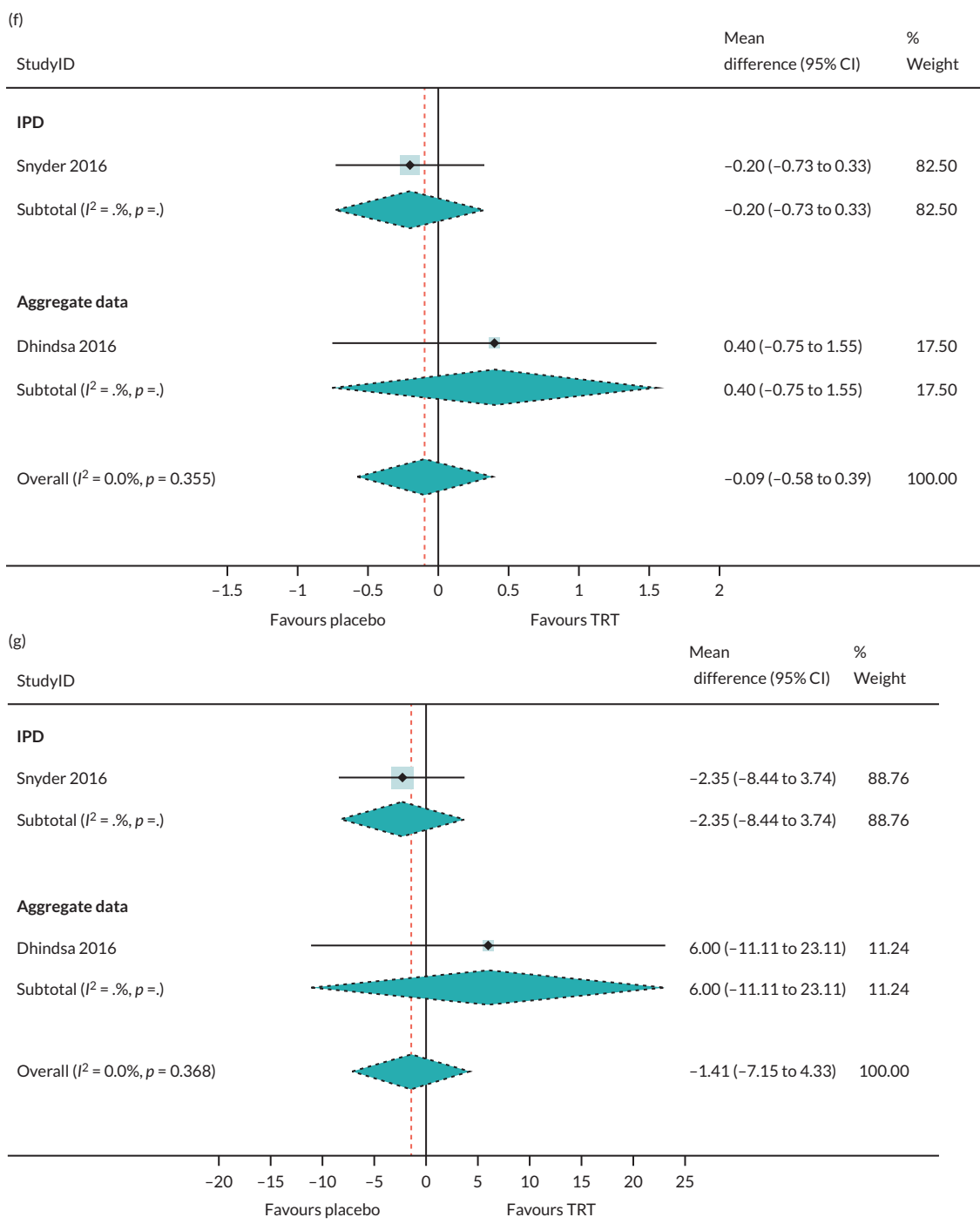


FIGURE 27 Two-stage meta-analysis for PDQ. (a) Positive mood; (b) negative mood; (c) sexual desire; (d) sexual enjoyment with partner; (e) sexual enjoyment without partner; (f) Satisfaction; (g) Percentage of full erection. a, Data presented as change from baseline.

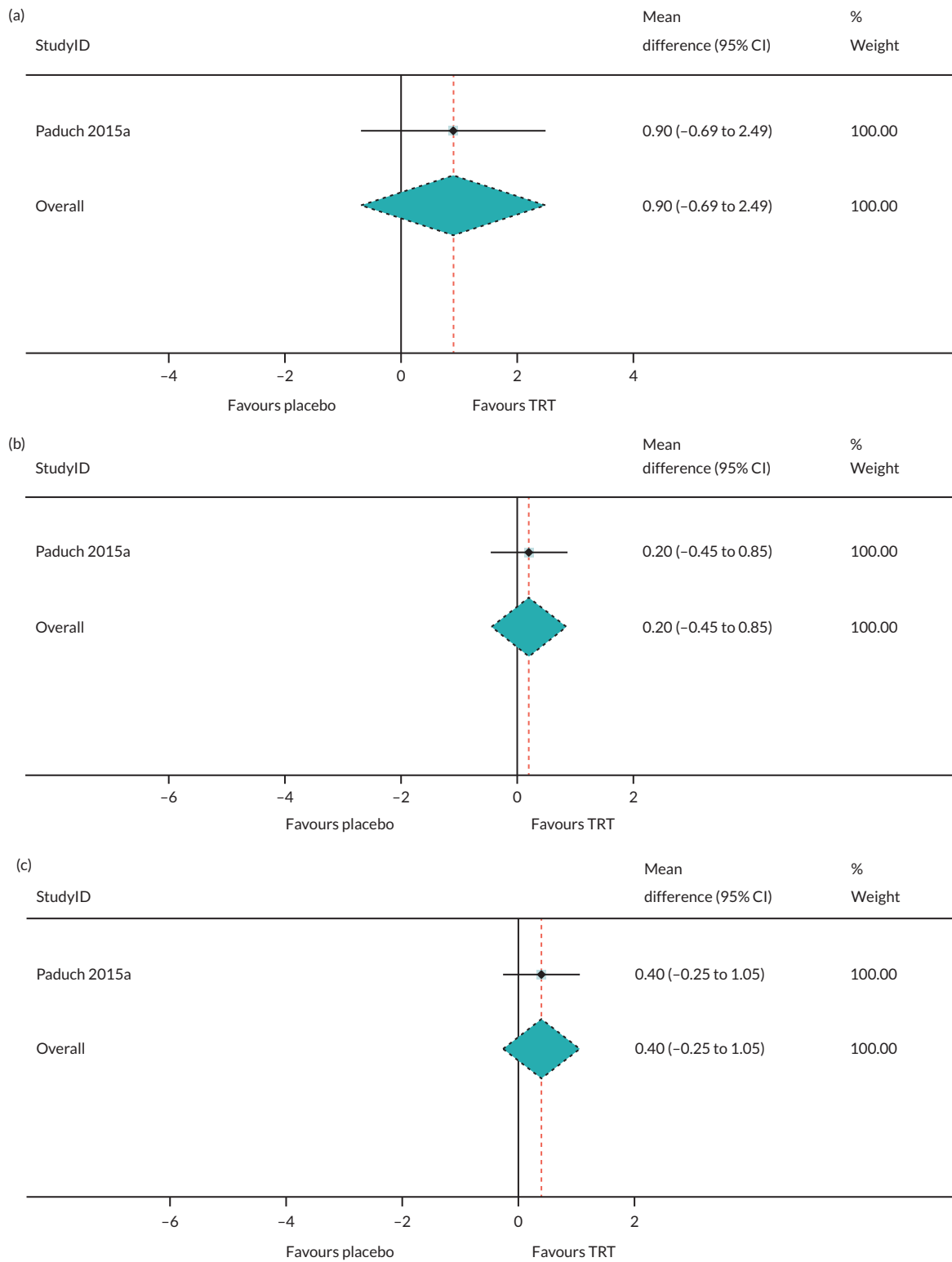


FIGURE 28 Two-stage meta-analysis for Male Sexual Health Questionnaire-Ejaculatory Dysfunction-Short Form. (a) Total score; (b) ejaculation volume; (c) frequency of ejaculation; (d) force of ejaculation; (e) bothered/satisfaction. a, Data presented as change from baseline. (continued)

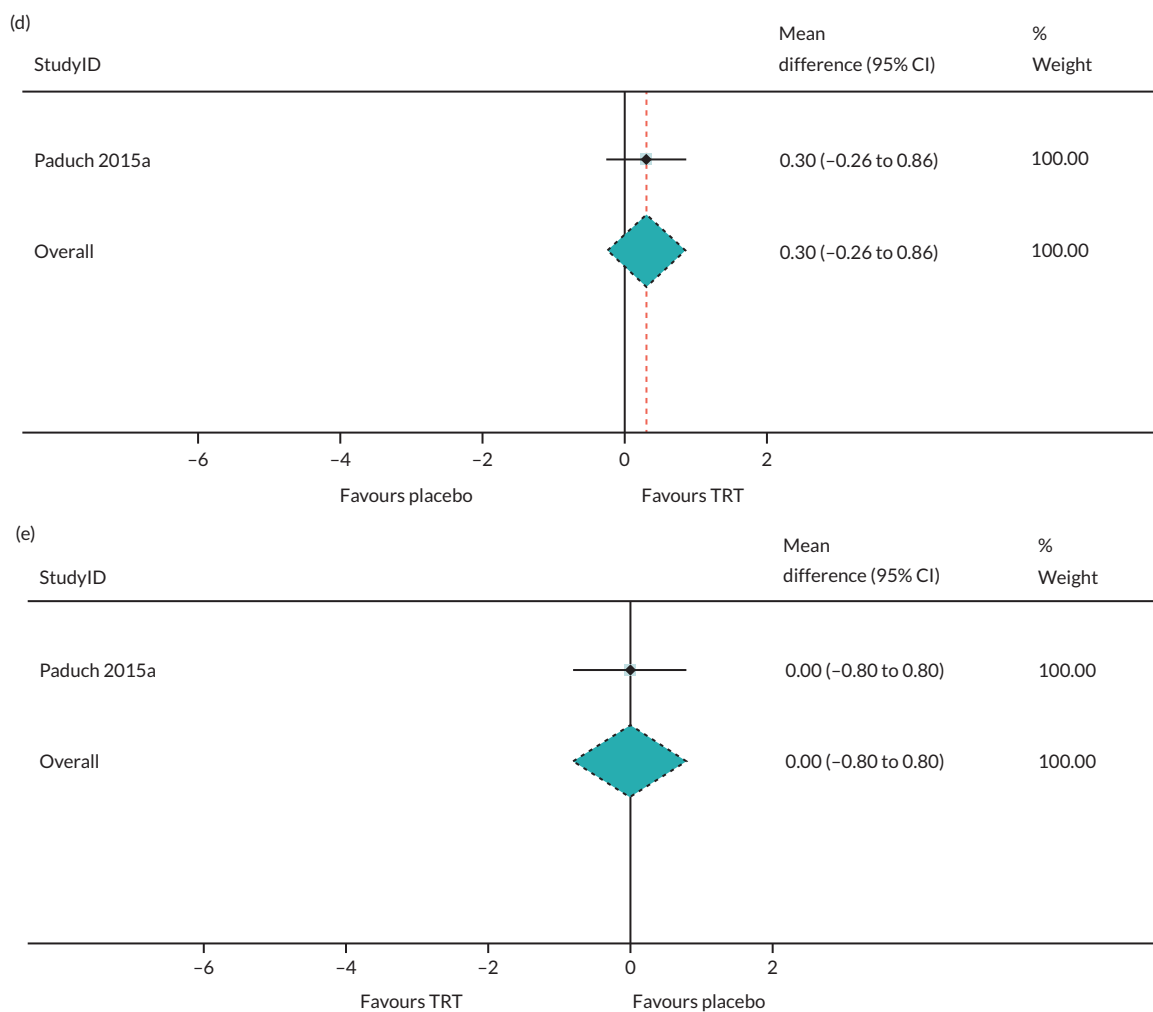


FIGURE 28 Two-stage meta-analysis for Male Sexual Health Questionnaire-Ejaculatory Dysfunction-Short Form. (a) Total score; (b) ejaculation volume; (c) frequency of ejaculation; (d) force of ejaculation; (e) bothered/satisfaction. a, Data presented as change from baseline.

TABLE 42 Subgroup analysis for IIEF-15 and its subscales

	Interaction (99% CI)	p-value
Age		
IIEF-15 Total	0.98 (0.56 to 1.71)	
IIEF-15 Erectile function	1.01 (0.81 to 1.26)	0.92
IIEF-15 Orgasmic function	1.00 (0.88 to 1.12)	0.94
IIEF-15 Sexual desire	1.01 (0.96 to 1.05)	0.77
IIEF-15 Intercourse satisfaction	0.99 (0.88 to 1.12)	0.89
IIEF-15 Overall satisfaction	0.99 (0.93 to 1.04)	0.51
Total serum testosterone		
IIEF-15 Total	0.98 (0.94 to 1.02)	
IIEF-15 Erectile function	0.86 (0.54 to 1.38)	0.41
IIEF-15 Orgasmic function	0.86 (0.69 to 1.07)	0.07
IIEF-15 Sexual desire	0.93 (0.82 to 1.06)	0.16
IIEF-15 Intercourse satisfaction	0.95 (0.75 to 1.22)	0.63
IIEF-15 Overall satisfaction	0.97 (0.83 to 1.12)	0.55

continued

TABLE 42 Subgroup analysis for IIEF-15 and its subscales (*continued*)

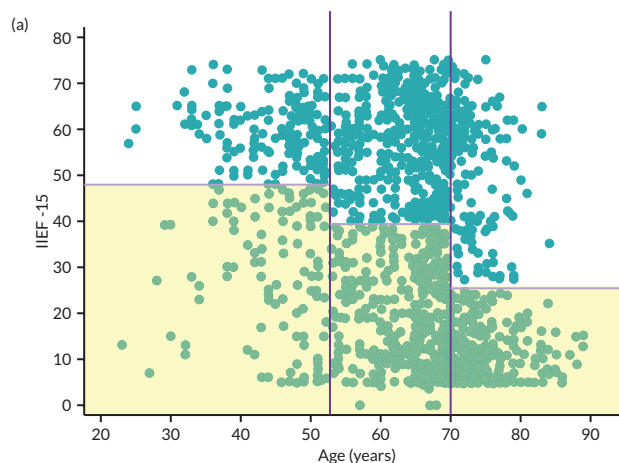
	Interaction (99% CI)	p-value
BMI		
AMS	0.89 (0.63 to 1.25)	0.37
IIEF-15 Total	0.74 (0.41 to 1.33)	0.19
IIEF-15 Erectile function	0.89 (0.67 to 1.20)	0.33
IIEF-15 Orgasmic function	1.00 (0.89 to 1.12)	0.93
IIEF-15 Sexual desire	0.96 (0.90 to 1.02)	0.08
IIEF-15 Intercourse satisfaction	0.93 (0.81 to 1.06)	0.14
IIEF-15 Overall satisfaction	0.96 (0.88 to 1.03)	0.13

TABLE 43 Sexual function during testosterone treatment in men classified according to identified thresholds

	Mean (SD); n	p-value
IIEF-15 at follow-up vs. age (years)		
< 52	46.8 (19.4); 237	<0.001
52–70	38.8 (21.9); 819	
> 70	26.8 (21.1); 345	
IIEF-15 at follow-up vs. baseline total serum testosterone (nmol/l)		
9.8	30.7 (22.1); 545	<0.001
≥ 9.8	42.2 (23.9); 306	
IIEF-15 at baseline vs. baseline total serum testosterone (nmol/l)		
< 9.8	26.9 (20.1); 605	<0.001
≥ 9.8	39.0 (22.5); 345	
IIEF-15 at follow-up vs. BMI (kg/m²)		
< 30.6	40.0 (22.2); 710	<0.001
≥ 30.6	34.4 (21.9); 702	

Note

Mean and standard deviation (SD) are shown for international inventory of erectile function 15 (IIEF-15) score.

**FIGURE 29** Thresholds for sexual function experienced during testosterone treatment in men. Scatterplots of IIEF-15 during testosterone vs. (a) patient age; (b) serum baseline TT; (c) BMI. Purple line – significant threshold; yellow line and shading – mean IIEF-15 levels above, below or between thresholds. (*continued*)

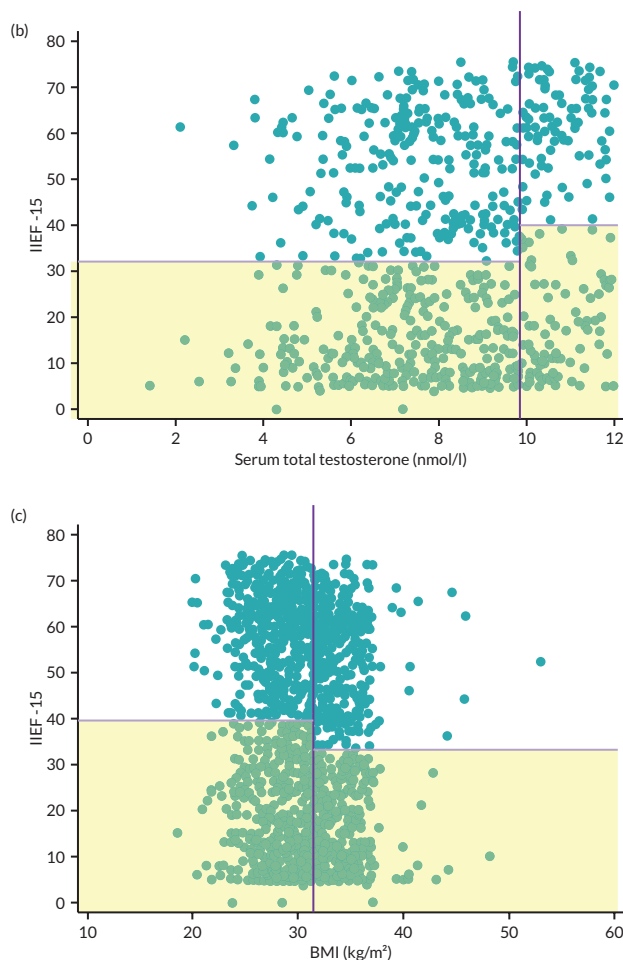


FIGURE 29 Thresholds for sexual function experienced during testosterone treatment in men. Scatterplots of IIEf-15 during testosterone vs. (a) patient age; (b) serum baseline TT; (c) BMI. Purple line – significant threshold; yellow line and shading – mean IIEF-15 levels above, below or between thresholds.

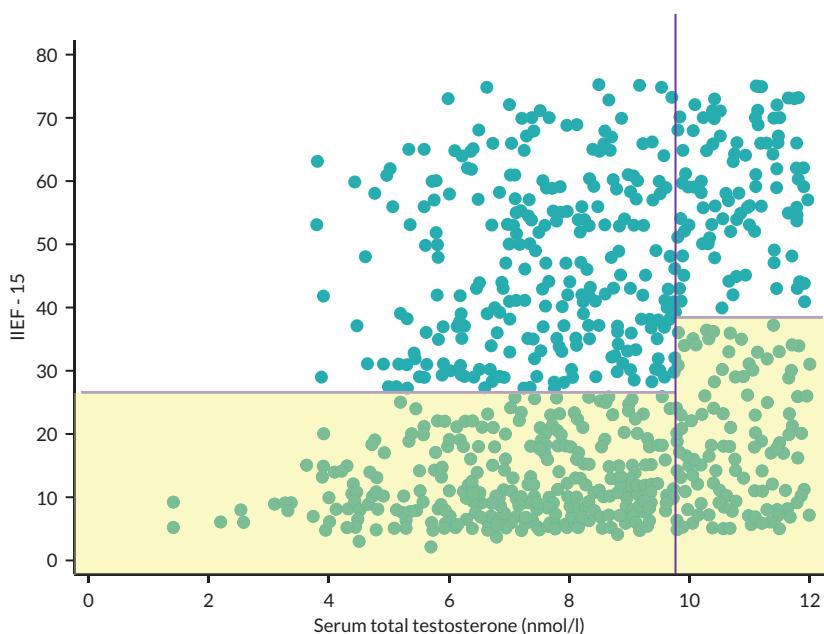


FIGURE 30 Scatterplot of baseline IIEf-15 vs. serum baseline TT. Purple line – significant threshold; yellow line and shading – mean IIEF-15 levels above, below or between thresholds. Only including participants with testosterone < 12 nmol/l.

TABLE 44 One-stage analysis for physiological marker outcomes

Outcome	Number of studies	TRT mean (SD), n	Placebo mean (SD), n	MD	95% CI	τ^2
Areal bone mineral density (g/cm ²)						
Total	4	1.21 (0.12); 352	1.20 (0.12); 312	0.00	(-0.00 to 0.01)	0.00
Subtotal	2	1.03 (0.10); 108	1.02 (0.09); 103	0.00	(-0.00 to 0.01)	0.00
Femoral neck	6	0.84 (0.14); 247	0.86 (0.16); 202	-0.00	(-0.01 to 0.01)	0.00
Lumbar spine	9	1.19 (0.21); 484	1.18 (0.21); 436	0.01	(0.00 to 0.02)	0.00
Thoracic spine	2	0.96 (0.17); 154	0.92 (0.13); 132	0.01	(-0.01 to 0.03)	0.00
Total hip	8	1.03 (0.15); 372	1.04 (0.15); 322	0.00	(-0.00 to 0.01)	0.00
Trochanter	3	0.79 (0.13); 86	0.76 (0.11); 62	0.00	(-0.01 to 0.01)	0.00
Intertrochanter	2	2.06 (5.55); 37	1.16 (0.16); 37	0.01	(-0.00 to 0.03)	0.00
Pelvis	2	1.23 (0.19); 148	1.19 (0.18); 116	0.01	(-0.00 to 0.02)	0.00
Left arm	3	0.85 (0.12); 175	0.83 (0.07); 151	-0.00	(-0.02 to 0.01)	0.00
Right arm	3	0.86 (0.13); 175	0.83 (0.08); 152	0.00	(-0.00 to 0.01)	0.00
Left plus right arm	1	1.64 (0.13); 20	1.58 (0.40); 19	-0.00	(-0.02 to 0.01)	
Left leg	3	1.26 (0.18); 166	1.23 (0.11); 135	0.00	(-0.01 to 0.01)	0.00
Right leg	3	1.26 (0.14); 162	1.25 (0.15); 132	0.01	(0.00 to 0.02)	0.00
Left plus right leg	1	2.45 (0.29); 20	2.31 (0.63); 18	0.02	(-0.03 to 0.07)	
Left rib	2	0.71 (0.09); 153	0.68 (0.08); 130	0.01	(-0.00 to 0.02)	0.00
Right rib	2	0.72 (0.09); 155	0.69 (0.07); 134	-0.00	(-0.01 to 0.01)	0.00
Head	2	2.13 (0.32); 156	2.05 (0.33); 134	0.00	(-0.02 to 0.02)	0.00
Shaft	1	1.21 (0.18); 47	1.15 (0.14); 23	0.00	(-0.01 to 0.02)	
Wards	1	0.61 (0.17); 47	0.56 (0.12); 23	-0.01	(-0.03 to 0.01)	
Volumetric bone mineral density (mg/cm ³)						
Spine trabecular	1	106.78 (32.37); 104	99.61 (27.07); 85	6.32	(4.54 to 8.09)	
Spine cortical	1	292.85 (43.05); 104	288.37 (43.83); 85	8.05	(5.83 to 10.28)	
Spine whole	1	199.57 (37.21); 104	194.61 (34.75); 85	7.86	(5.98 to 9.74)	
Hips trabecular	1	187.12 (35.03); 99	181.87 (32.92); 79	2.36	(1.41 to 3.31)	
Hips cortical	1	402.83 (46.10); 99	395.65 (47.97); 79	3.52	(1.52 to 5.52)	
Hips whole	1	251.22 (38.56); 09	245.19 (37.80); 79	2.84	(1.71 to 3.97)	

Notes

Values are mean (standard deviation); numbers.;

A fixed-effects analysis was used as only one study reported the outcome.

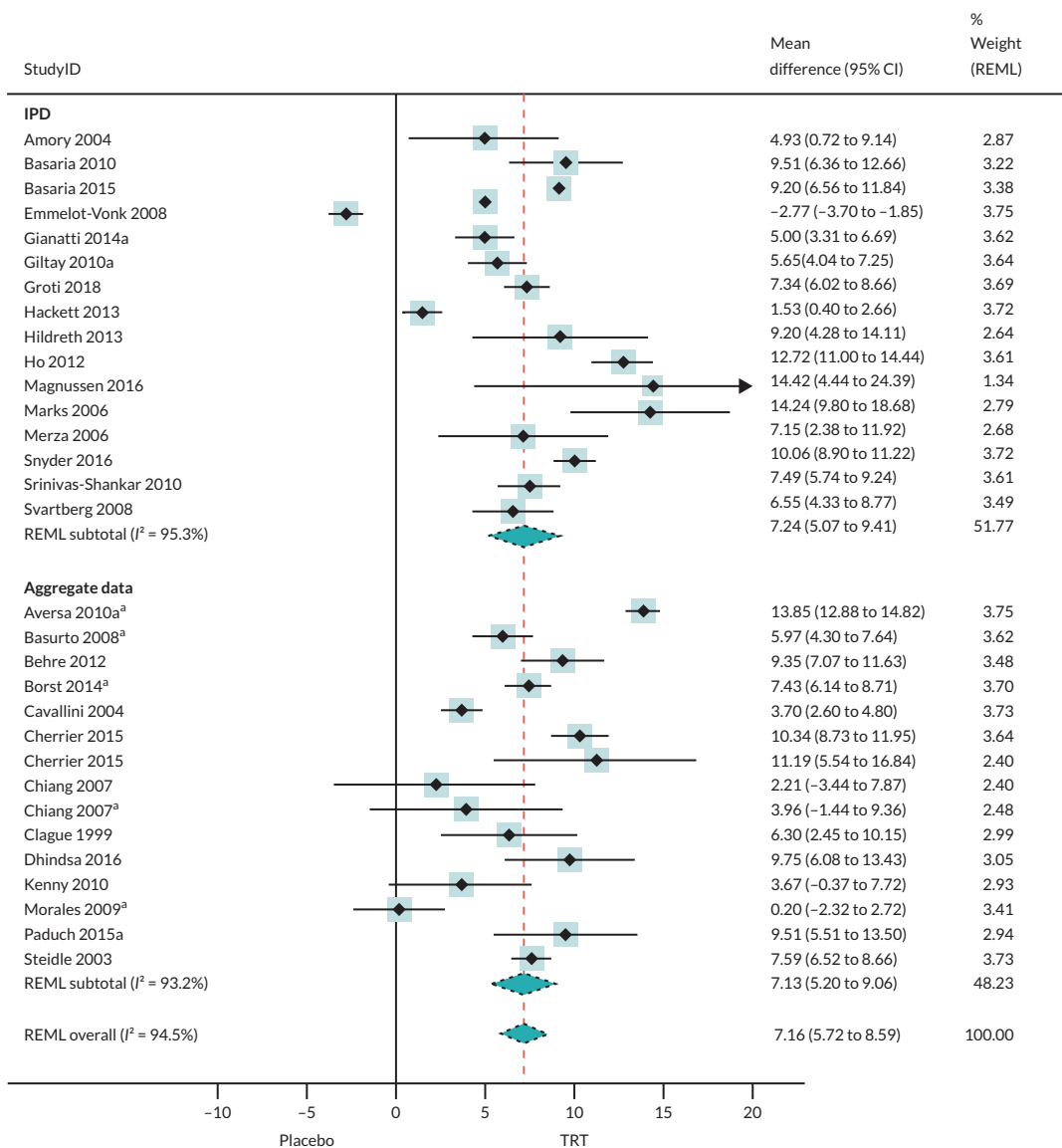


FIGURE 31 Two-stage meta-analysis for increase in testosterone (nmol/l). a, Data presented as change from baseline.

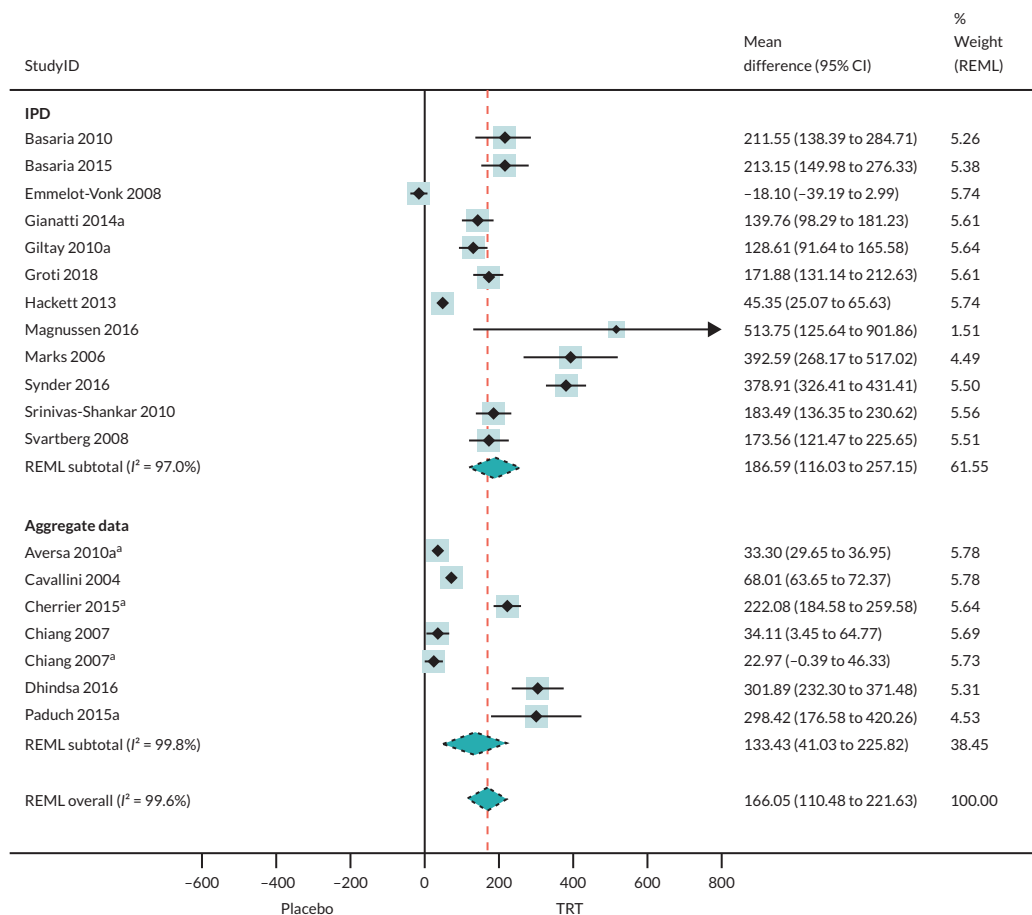


FIGURE 32 Two-stage meta-analysis for increase in free testosterone (pmol/l). a, Data presented as change from baseline.

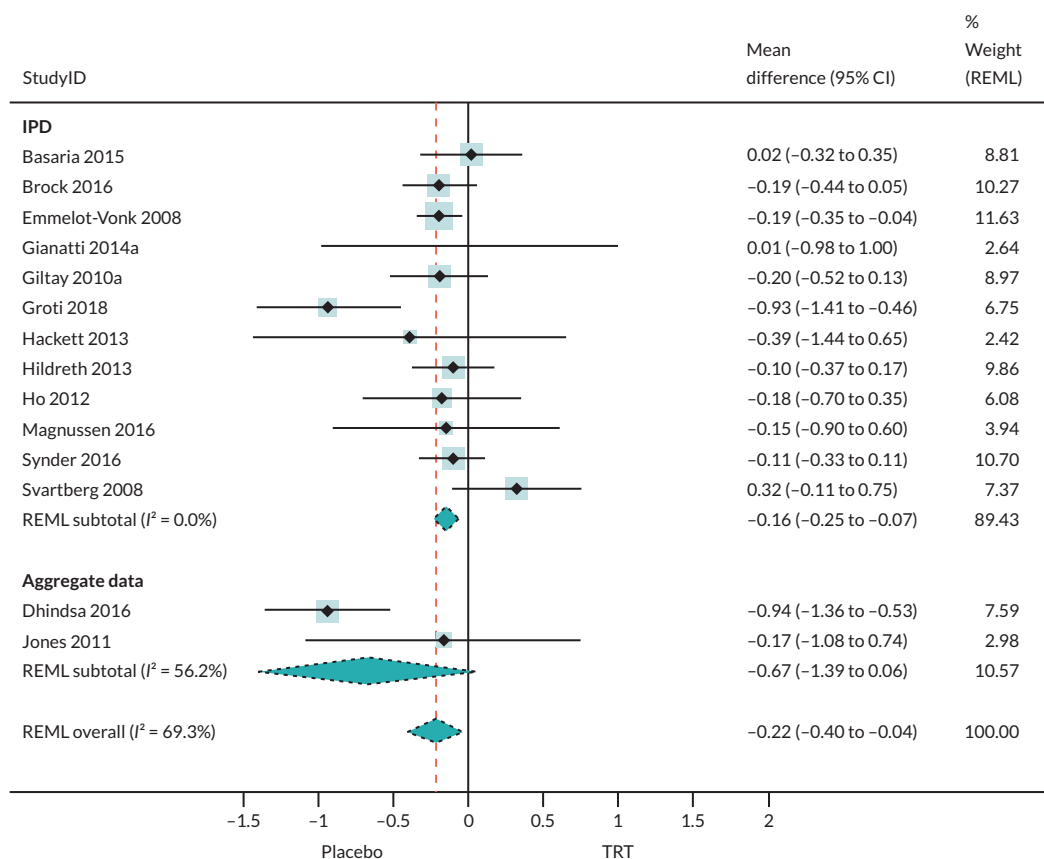


FIGURE 33 Two-stage meta-analysis for increase in fasting glucose (mmol/l).

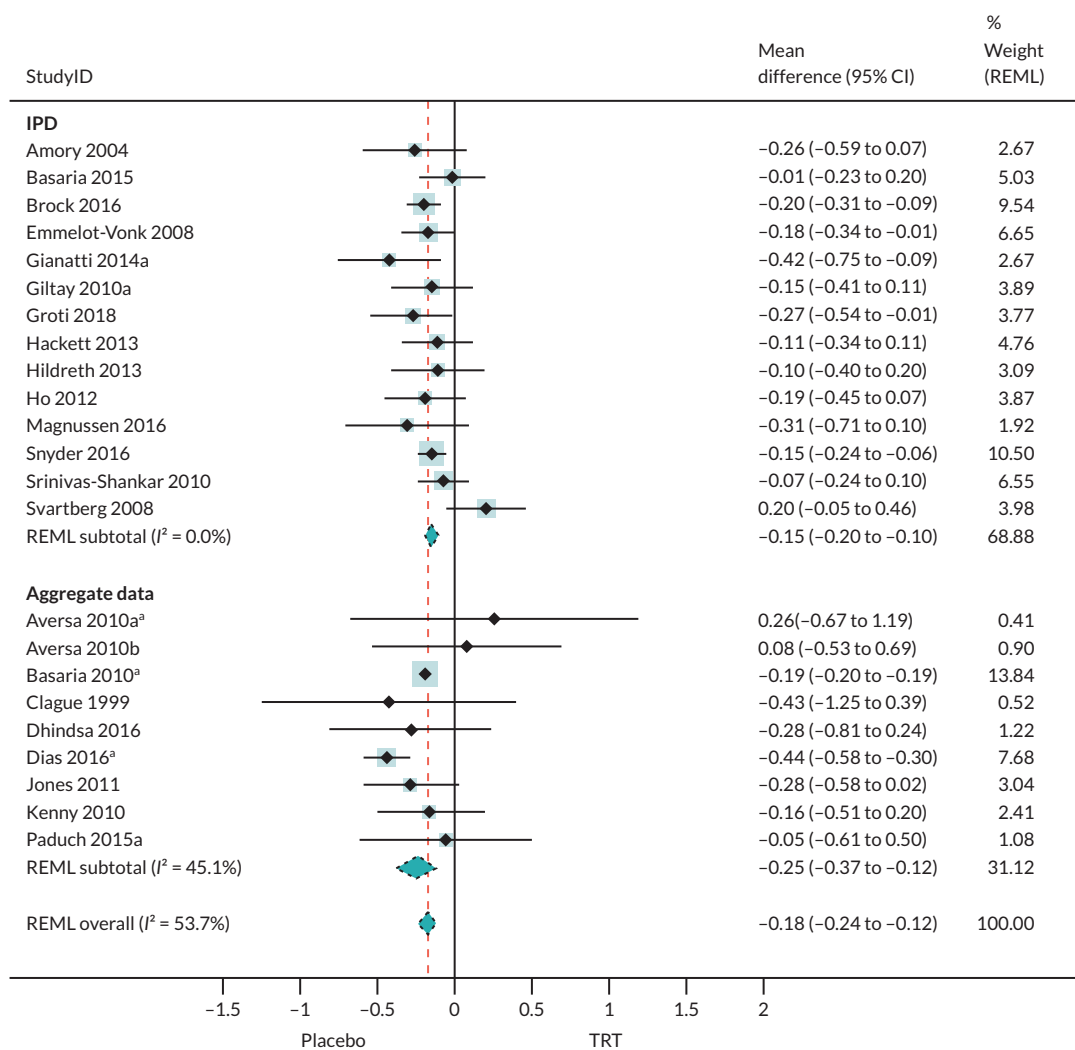


FIGURE 34 Two-stage meta-analysis for increase in cholesterol (mmol/l). a, Data presented as change from baseline.

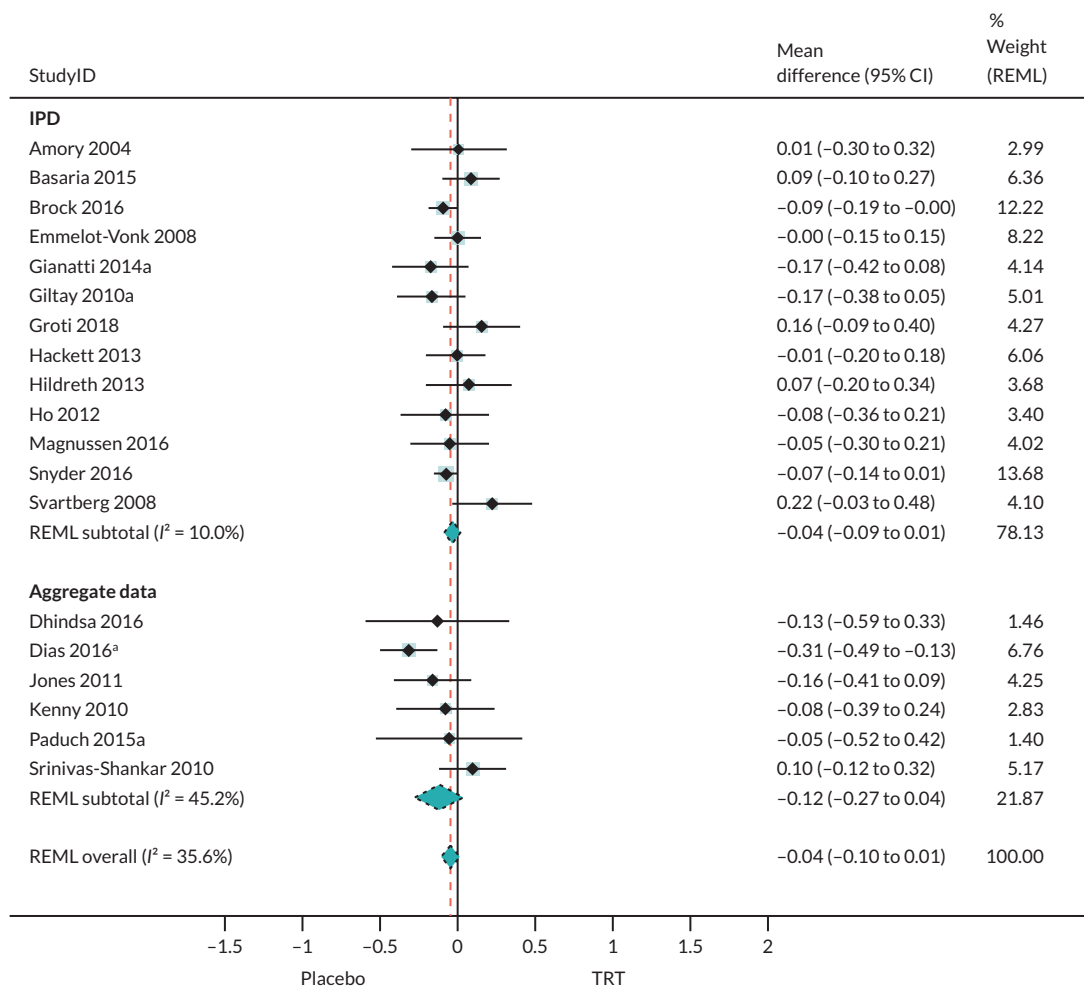


FIGURE 35 Two-stage meta-analysis for increase in low-density lipoprotein cholesterol (mmol/l). a, Data presented as change from baseline.

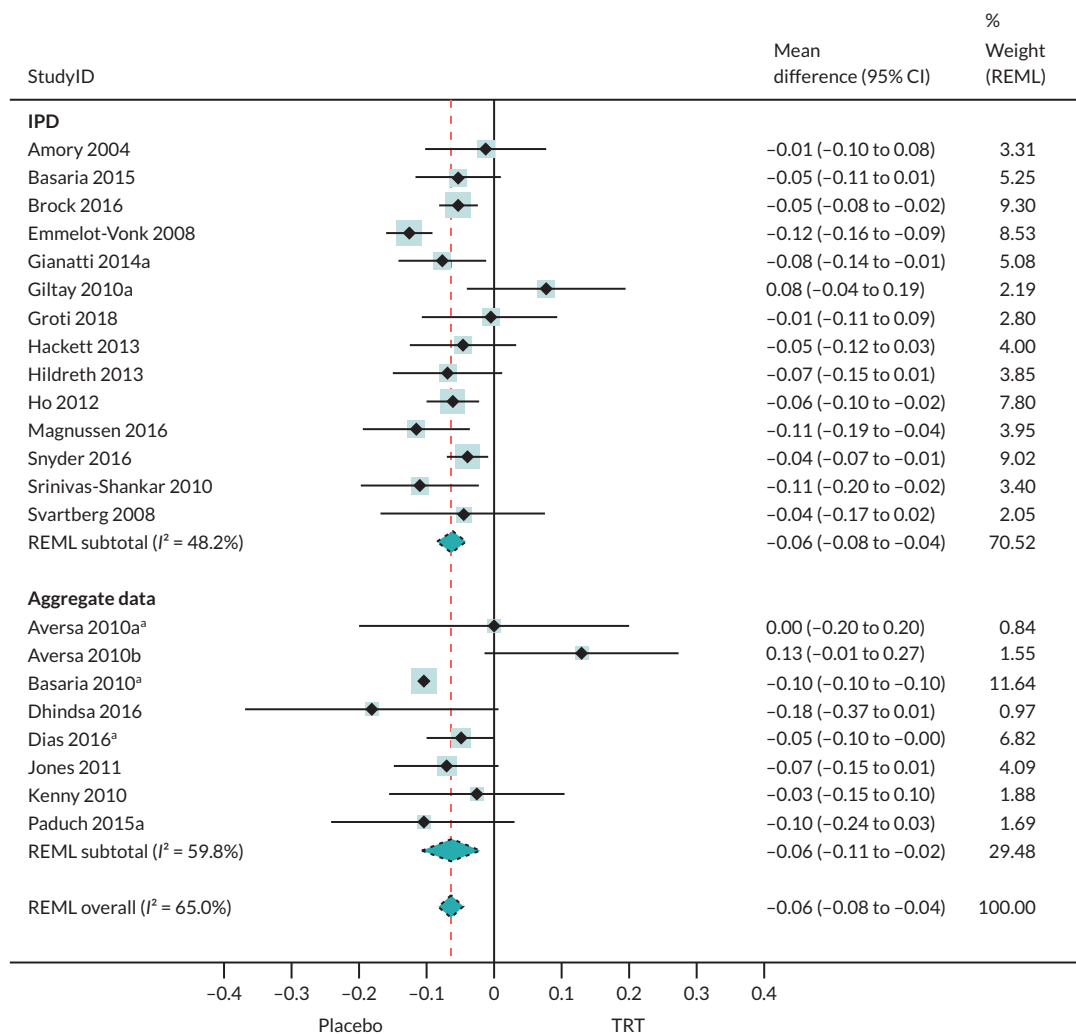


FIGURE 36 Two-stage meta-analysis for increase in high-density lipoprotein cholesterol (mmol/l). a, Data presented as change from baseline.

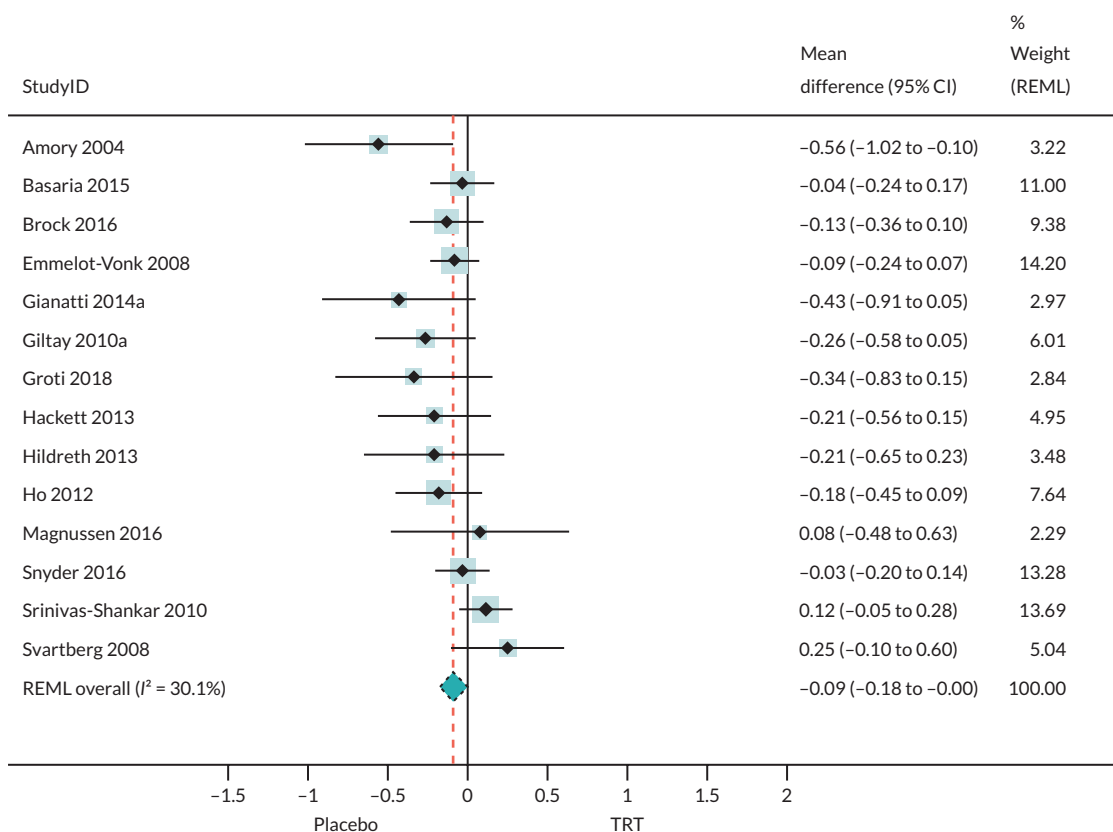


FIGURE 37 Two-stage meta-analysis for increase in triglycerides (mmol/l).

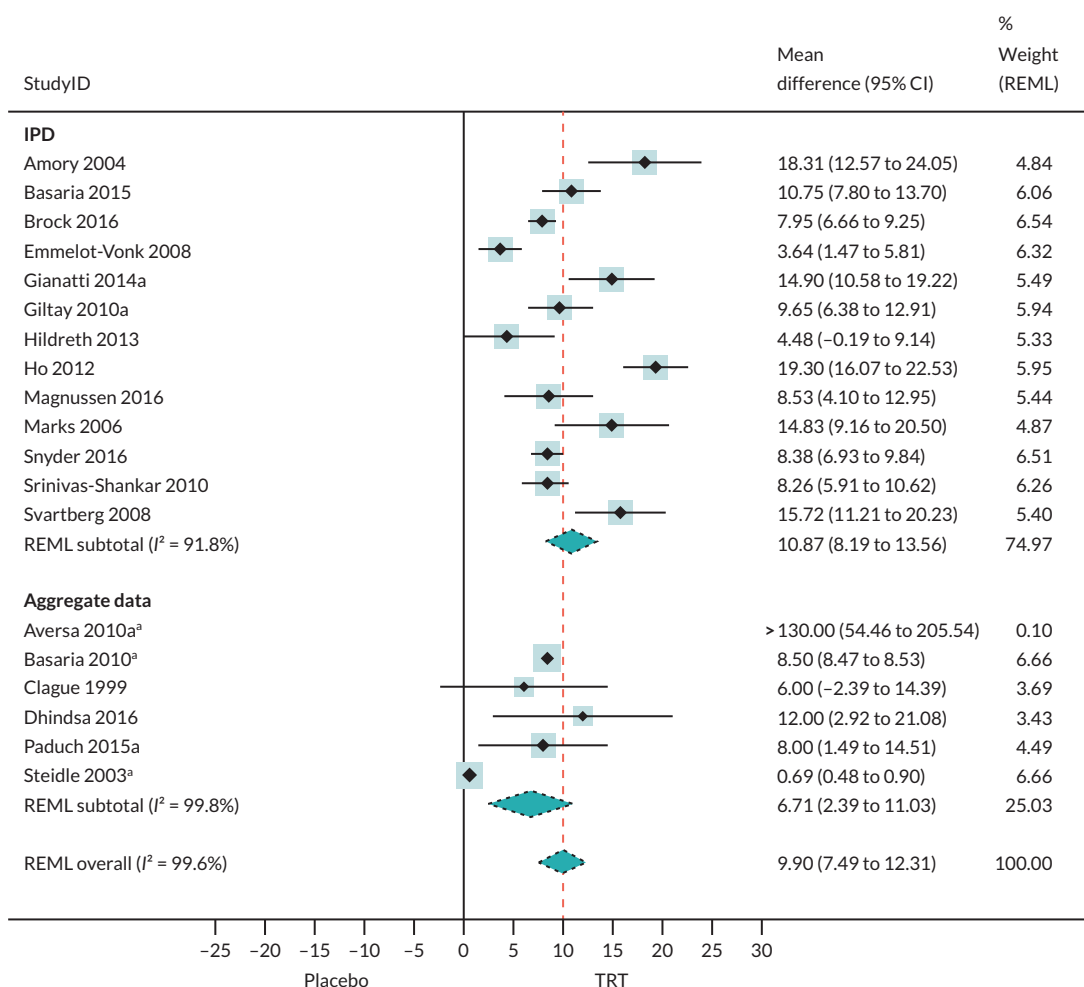


FIGURE 38 Two-stage meta-analysis for increase in Hb (g/l). a, Data presented as change from baseline.

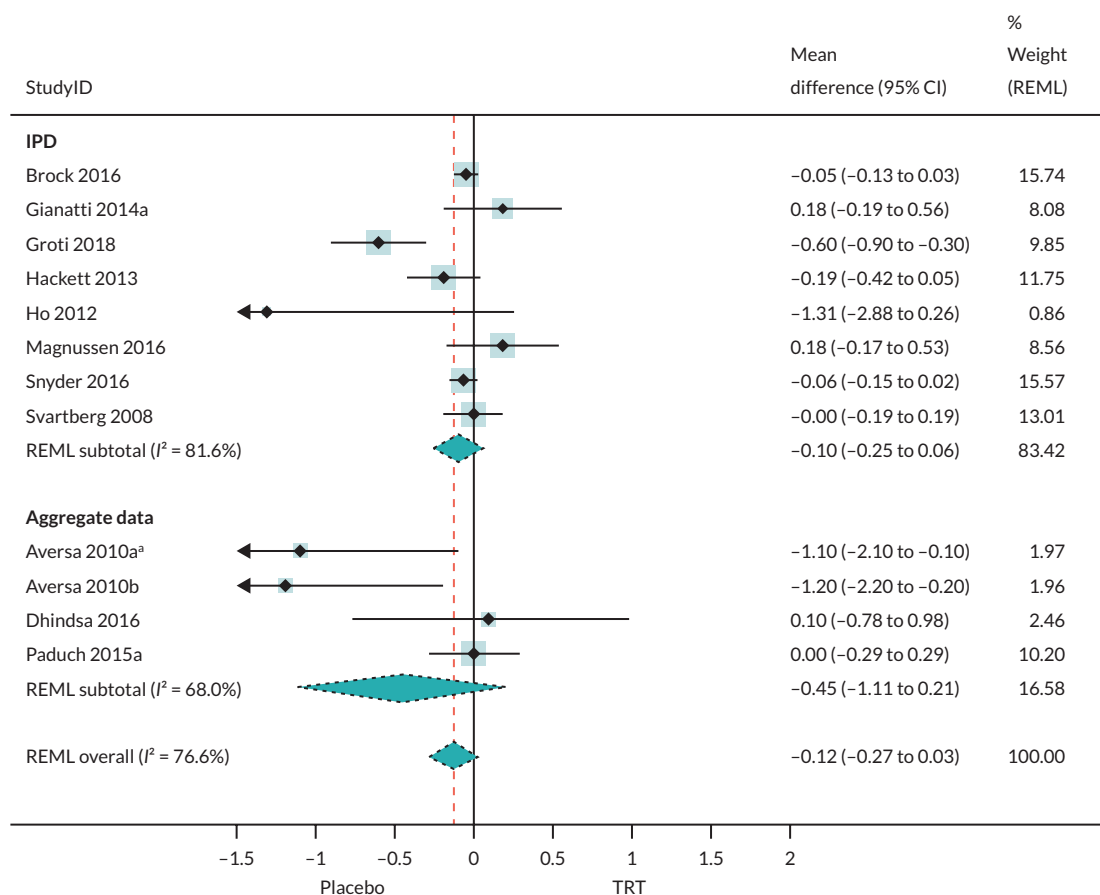


FIGURE 39 Two-stage meta-analysis for increase in HbA1c (%). a, Data presented as change from baseline.

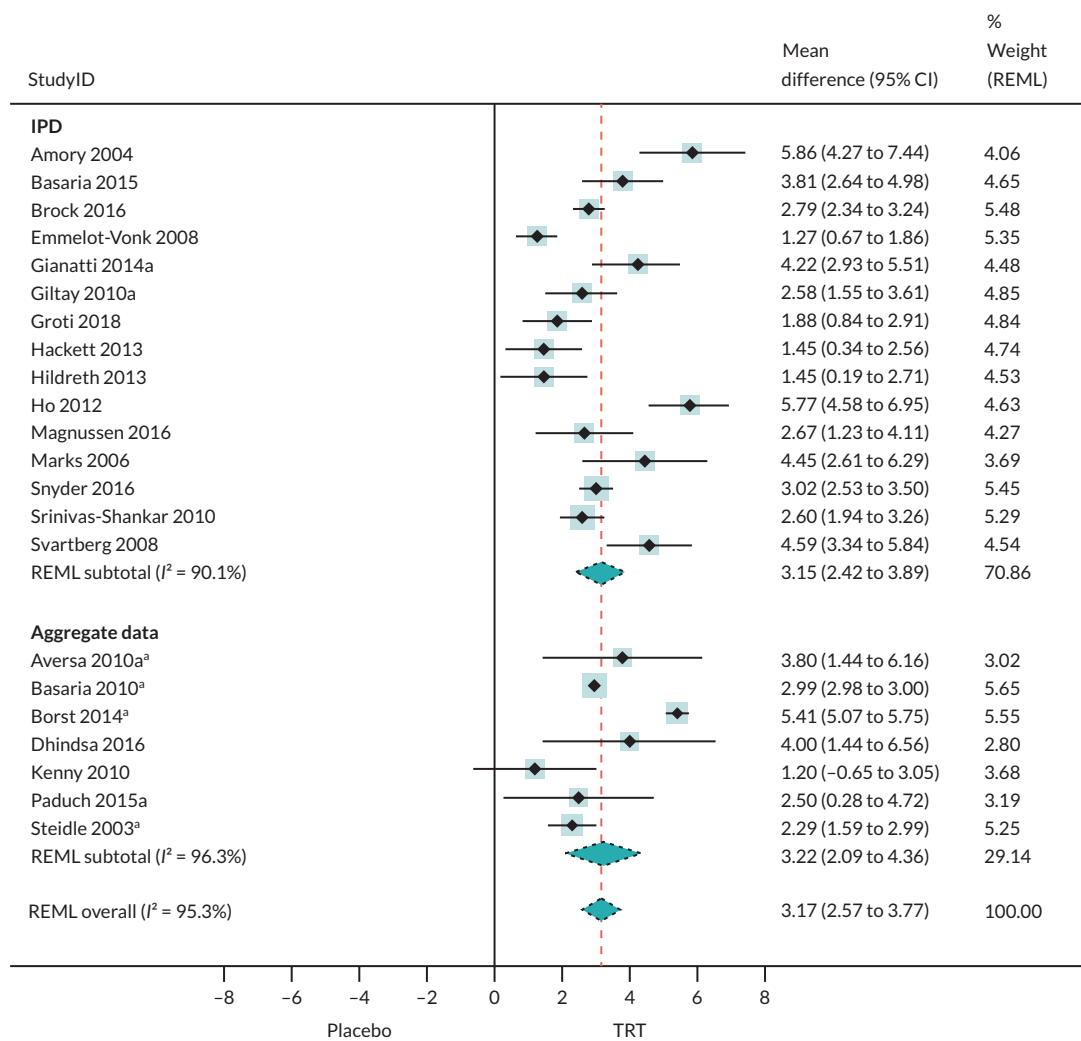


FIGURE 40 Two-stage meta-analysis for increase in haematocrit (%). a, Data presented as change from baseline.

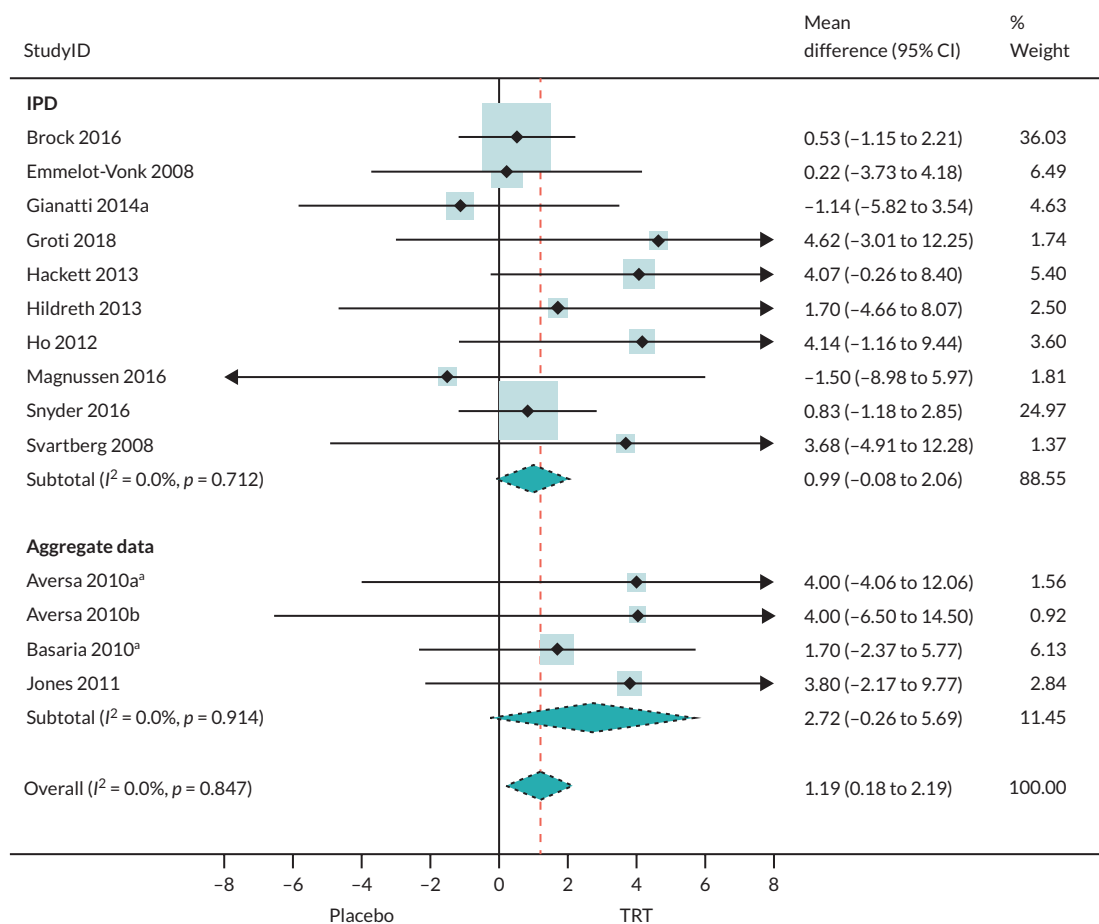


FIGURE 41 Two-stage meta-analysis for increase in SBP (mmHg). a, Data presented as change from baseline.

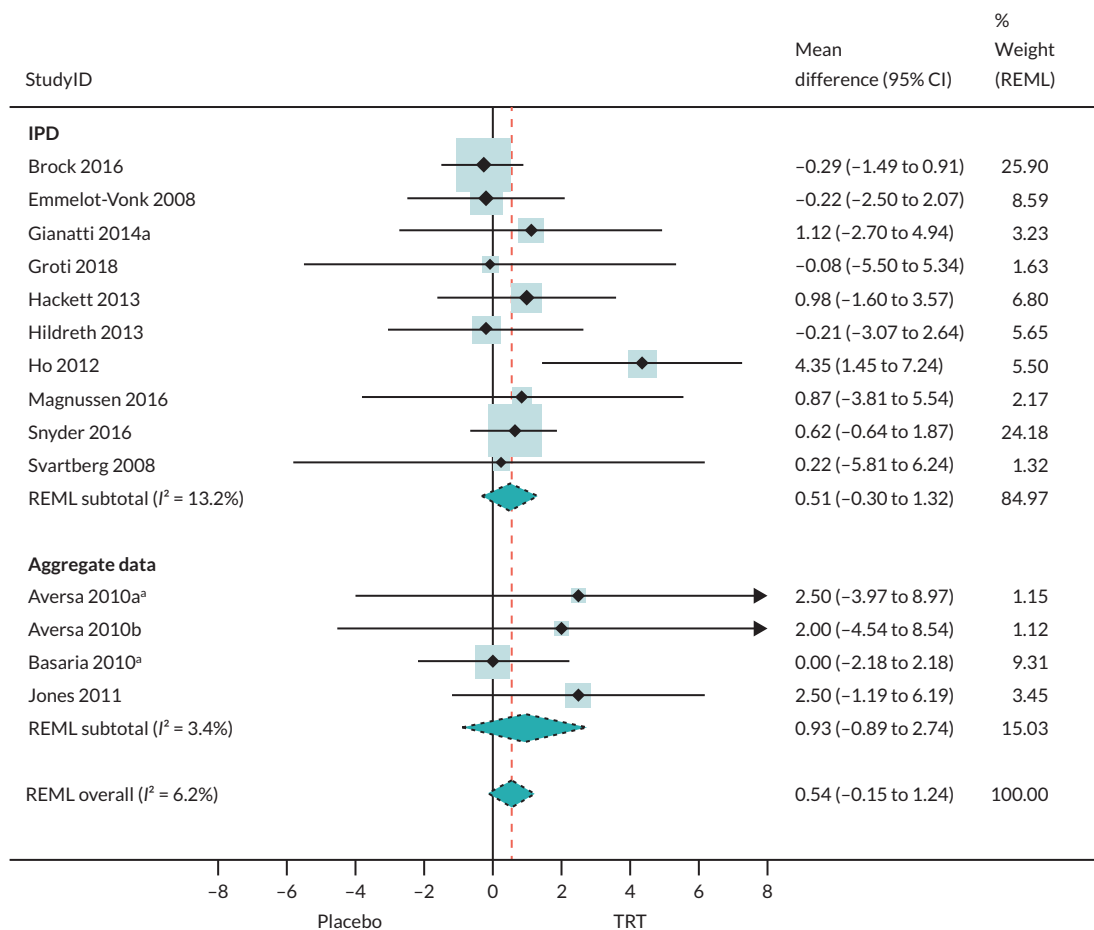


FIGURE 42 Two-stage meta-analysis for increase in DBP (mmHg) (nmol/l). a, Data presented as change from baseline.

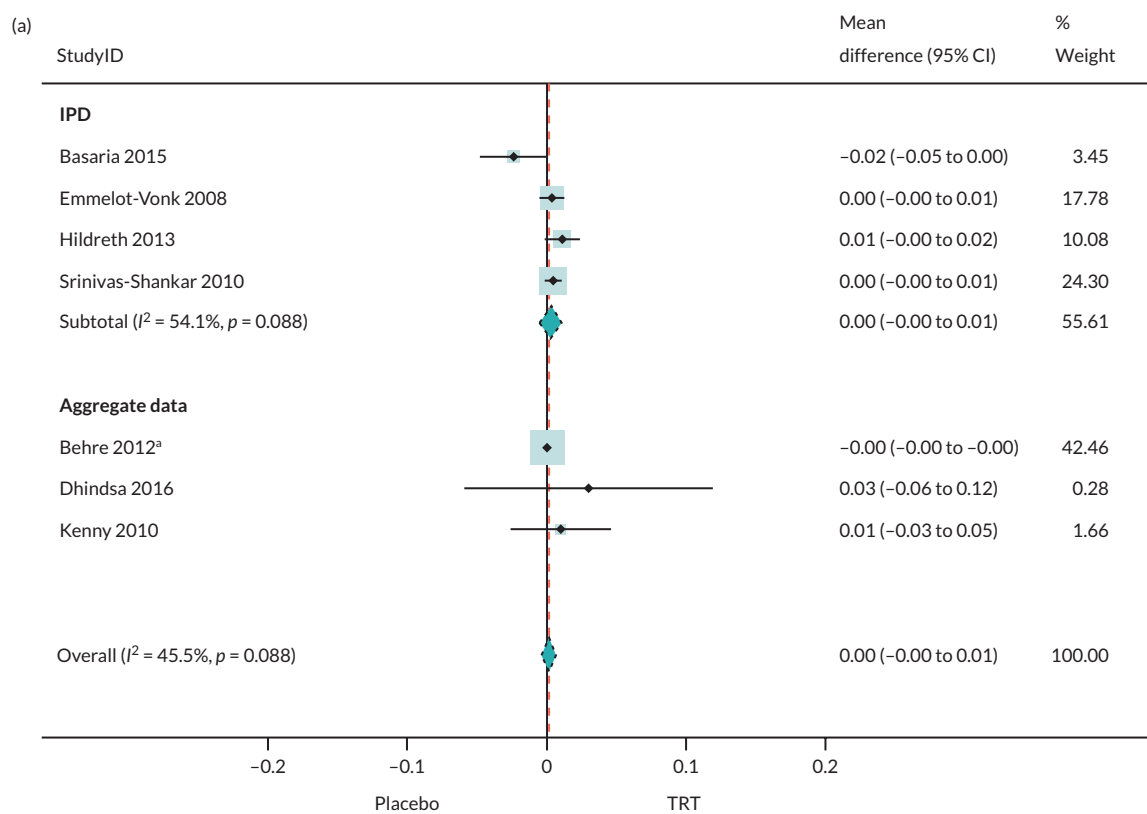


FIGURE 43 Two-stage meta-analysis for increase in bone mineral density (g/cm²). (a) Total score; (b) subtotal; (c) femoral neck; (d) lumbar spine; (e) thoracic spine; (f) total hip; (g) trochanter; (h) intertrochanter; (i) pelvis; (j) left arm; (k) right arm; (l) left leg; (m) right leg; (n) left rib; (o) right rib. a, Data presented as change from baseline. (continued)

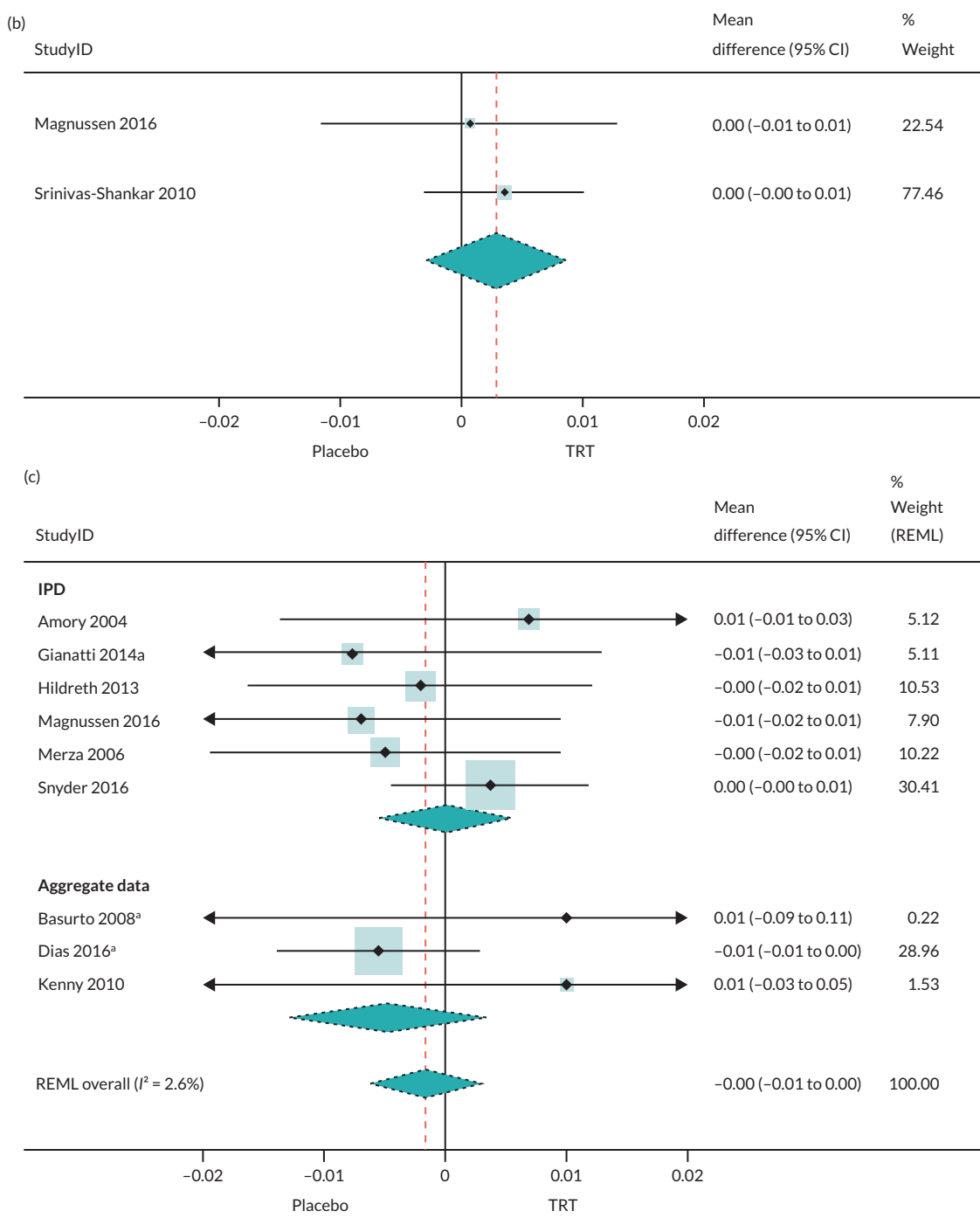


FIGURE 43 Two-stage meta-analysis for increase in bone mineral density (g/cm²). (a) Total score; (b) subtotal; (c) femoral neck; (d) lumbar spine; (e) thoracic spine; (f) total hip; (g) trochanter; (h) intertrochanter; (i) pelvis; (j) left arm; (k) right arm; (l) left leg; (m) right leg; (n) left rib; (o) right rib. a, Data presented as change from baseline. (continued)

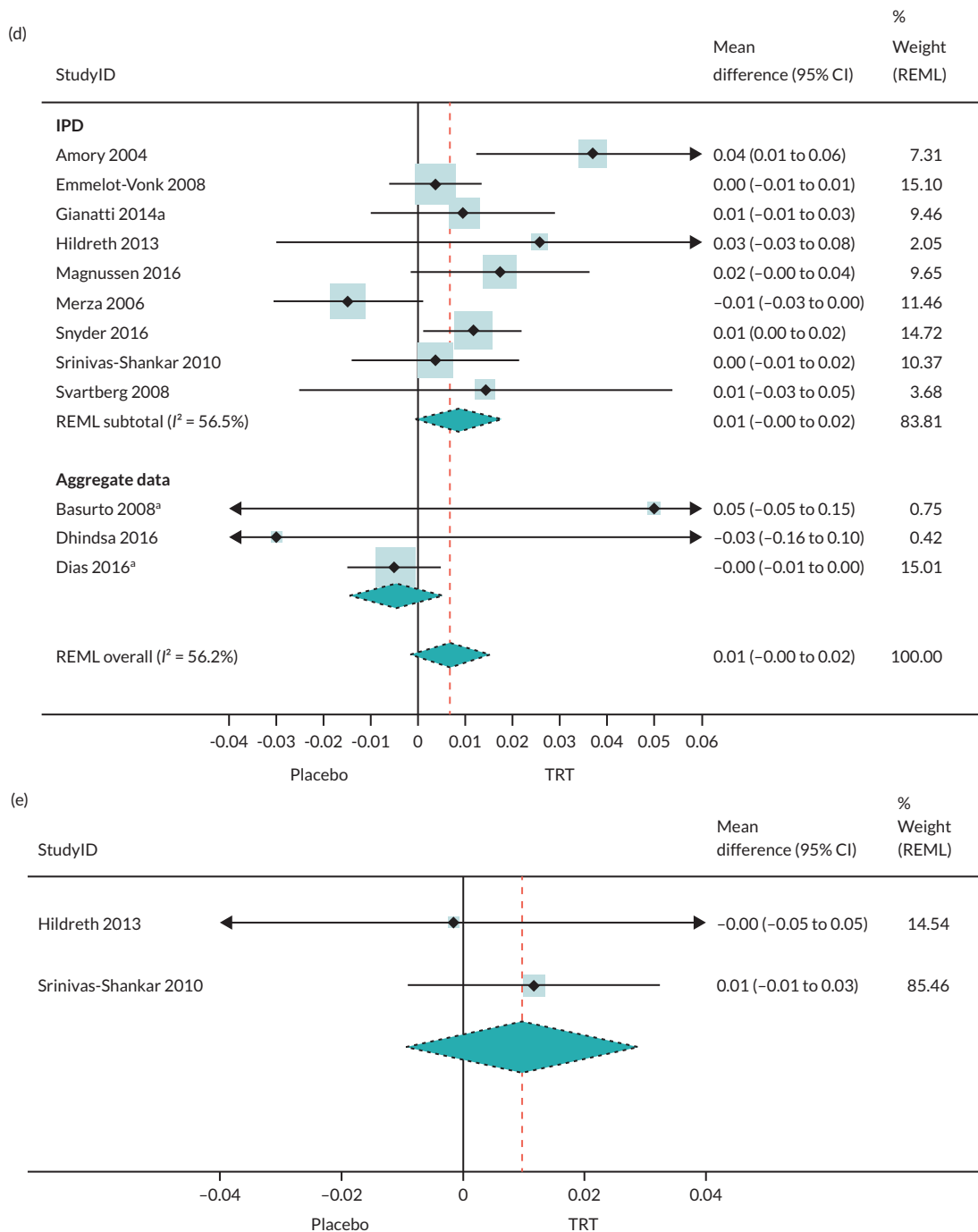


FIGURE 43 Two-stage meta-analysis for increase in bone mineral density (g/cm^2). (a) Total score; (b) subtotal; (c) femoral neck; (d) lumbar spine; (e) thoracic spine; (f) total hip; (g) trochanter; (h) intertrochanter; (i) pelvis; (j) left arm; (k) right arm; (l) left leg; (m) right leg; (n) left rib; (o) right rib. a, Data presented as change from baseline. (continued)

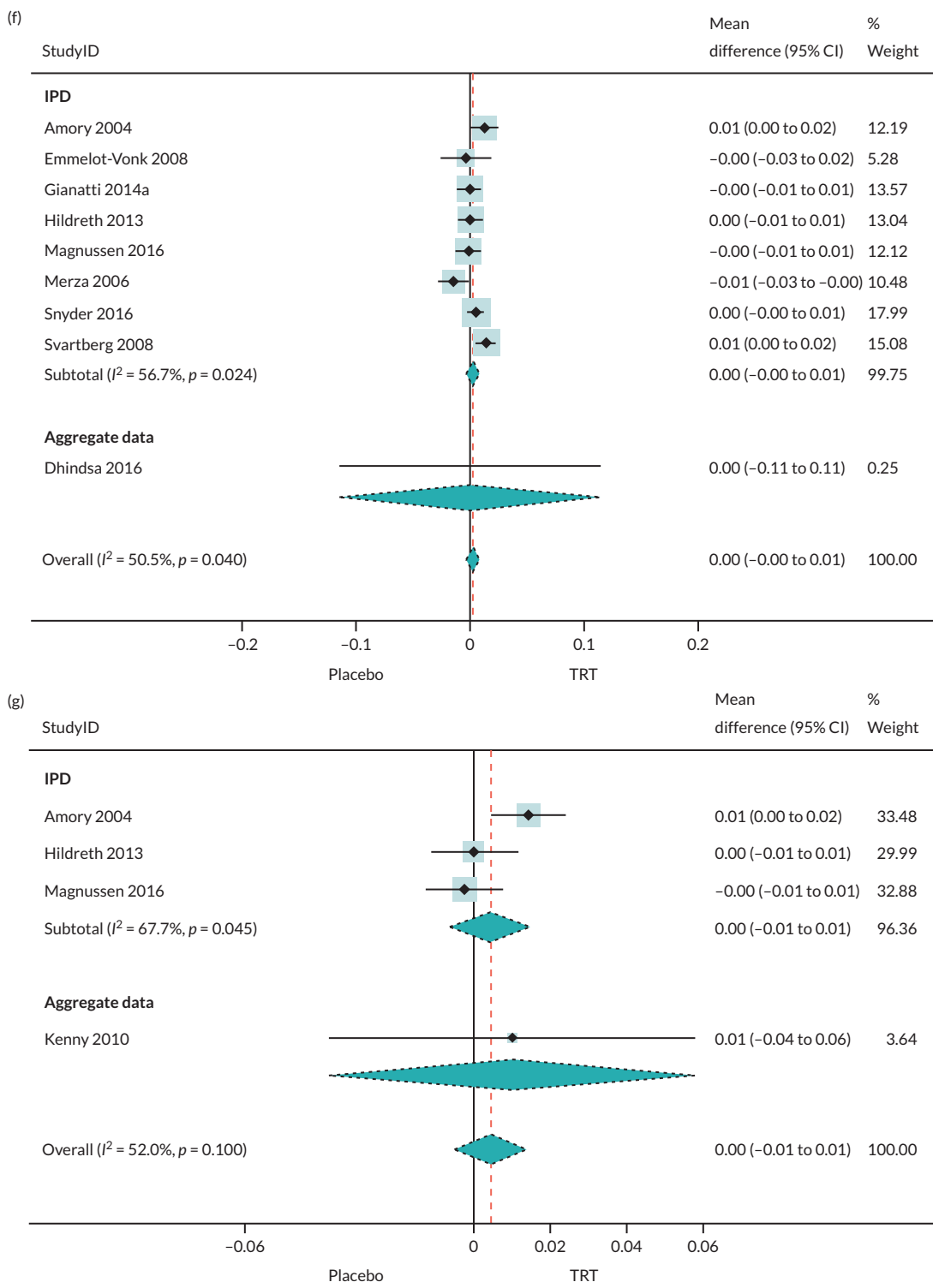


FIGURE 43 Two-stage meta-analysis for increase in bone mineral density (g/cm^2). (a) Total score; (b) subtotal; (c) femoral neck; (d) lumbar spine; (e) thoracic spine; (f) total hip; (g) trochanter; (h) intertrochanter; (i) pelvis; (j) left arm; (k) right arm; (l) left leg; (m) right leg; (n) left rib; (o) right rib. a, Data presented as change from baseline. (continued)

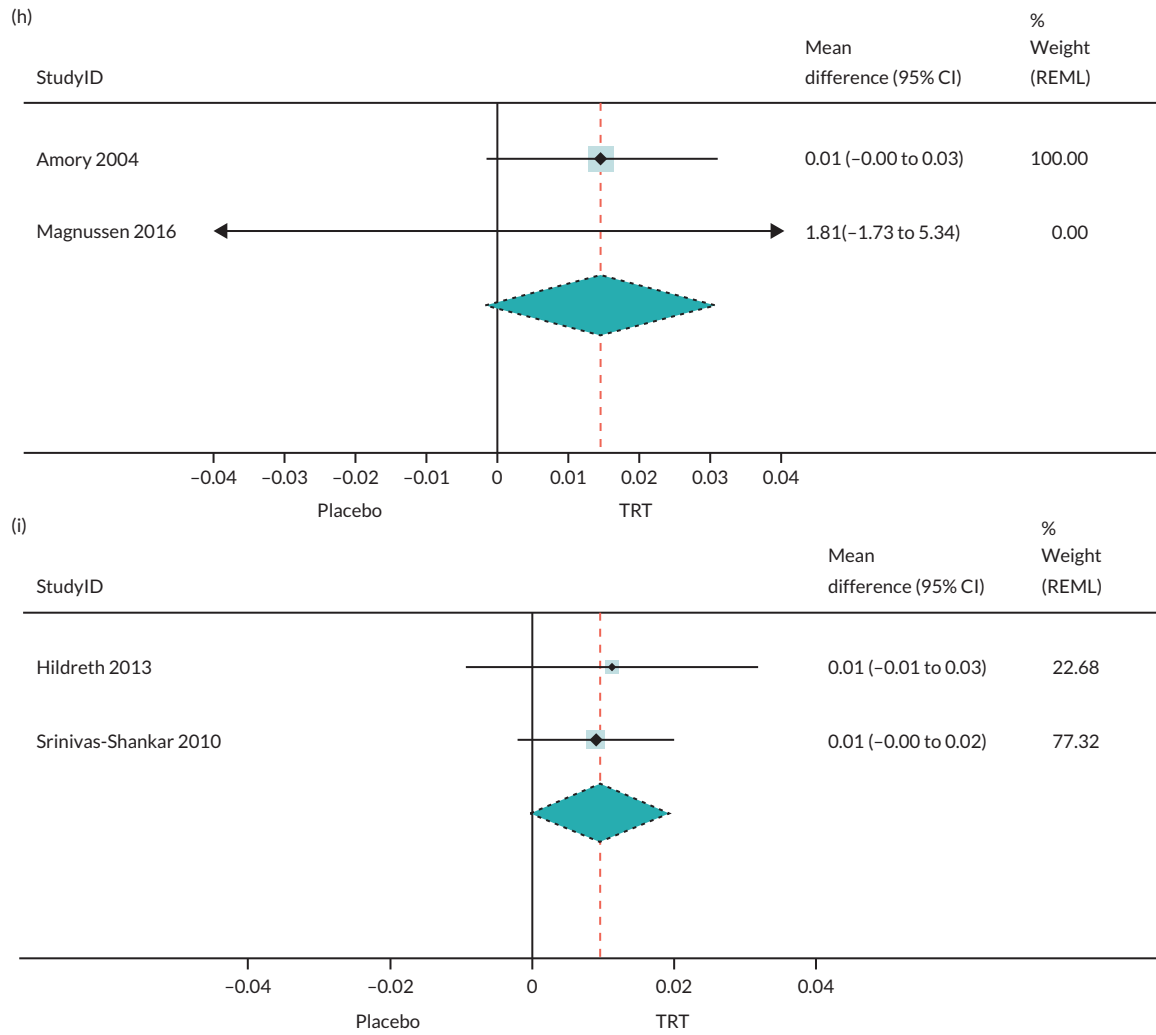


FIGURE 43 Two-stage meta-analysis for increase in bone mineral density (g/cm²). (a) Total score; (b) subtotal; (c) femoral neck; (d) lumbar spine; (e) thoracic spine; (f) total hip; (g) trochanter; (h) intertrochanter; (i) pelvis; (j) left arm; (k) right arm; (l) left leg; (m) right leg; (n) left rib; (o) right rib. a, Data presented as change from baseline. (continued)

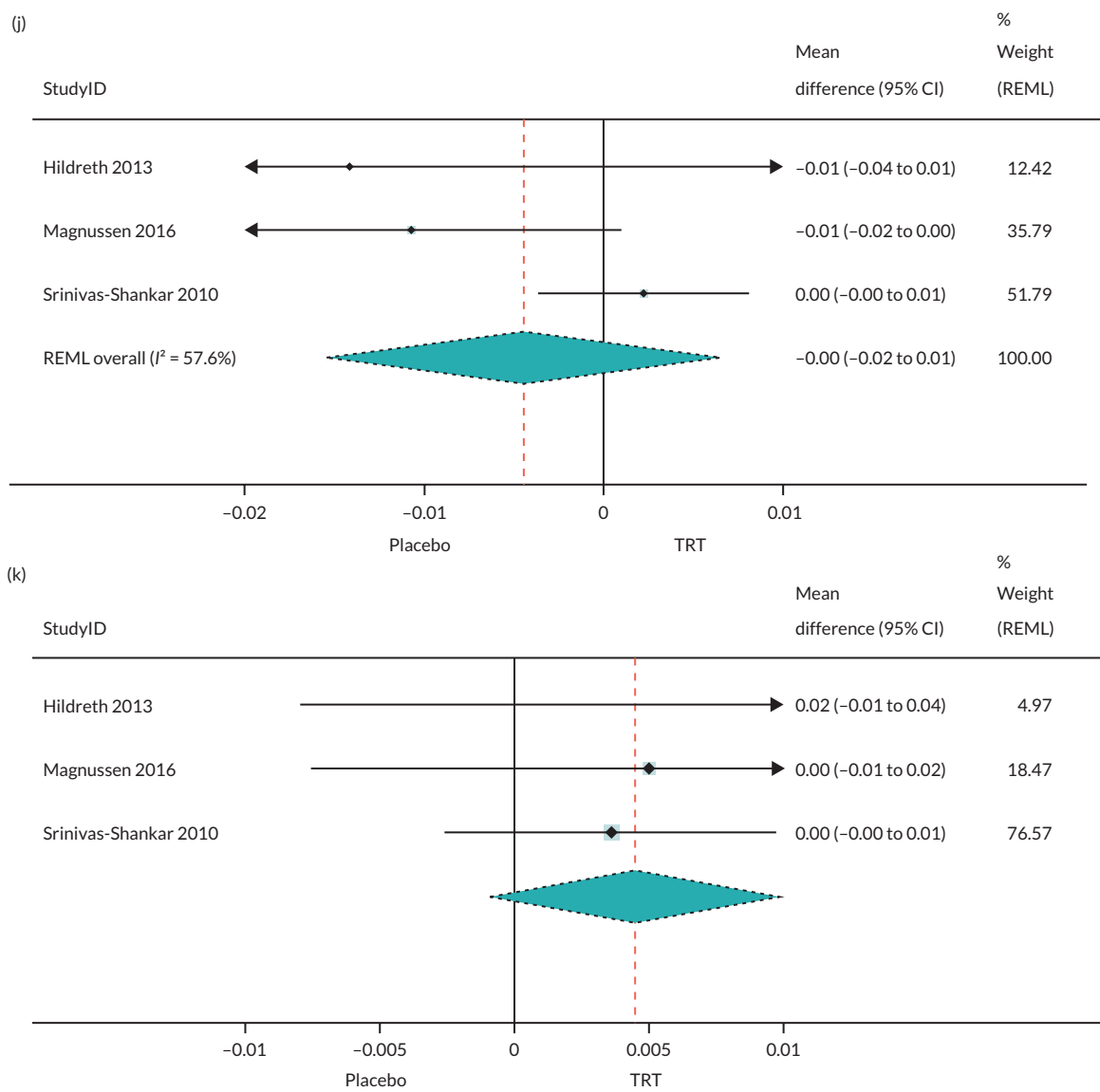


FIGURE 43 Two-stage meta-analysis for increase in bone mineral density (g/cm^2). (a) Total score; (b) subtotal; (c) femoral neck; (d) lumbar spine; (e) thoracic spine; (f) total hip; (g) trochanter; (h) intertrochanter; (i) pelvis; (j) left arm; (k) right arm; (l) left leg; (m) right leg; (n) left rib; (o) right rib. a, Data presented as change from baseline. (continued)

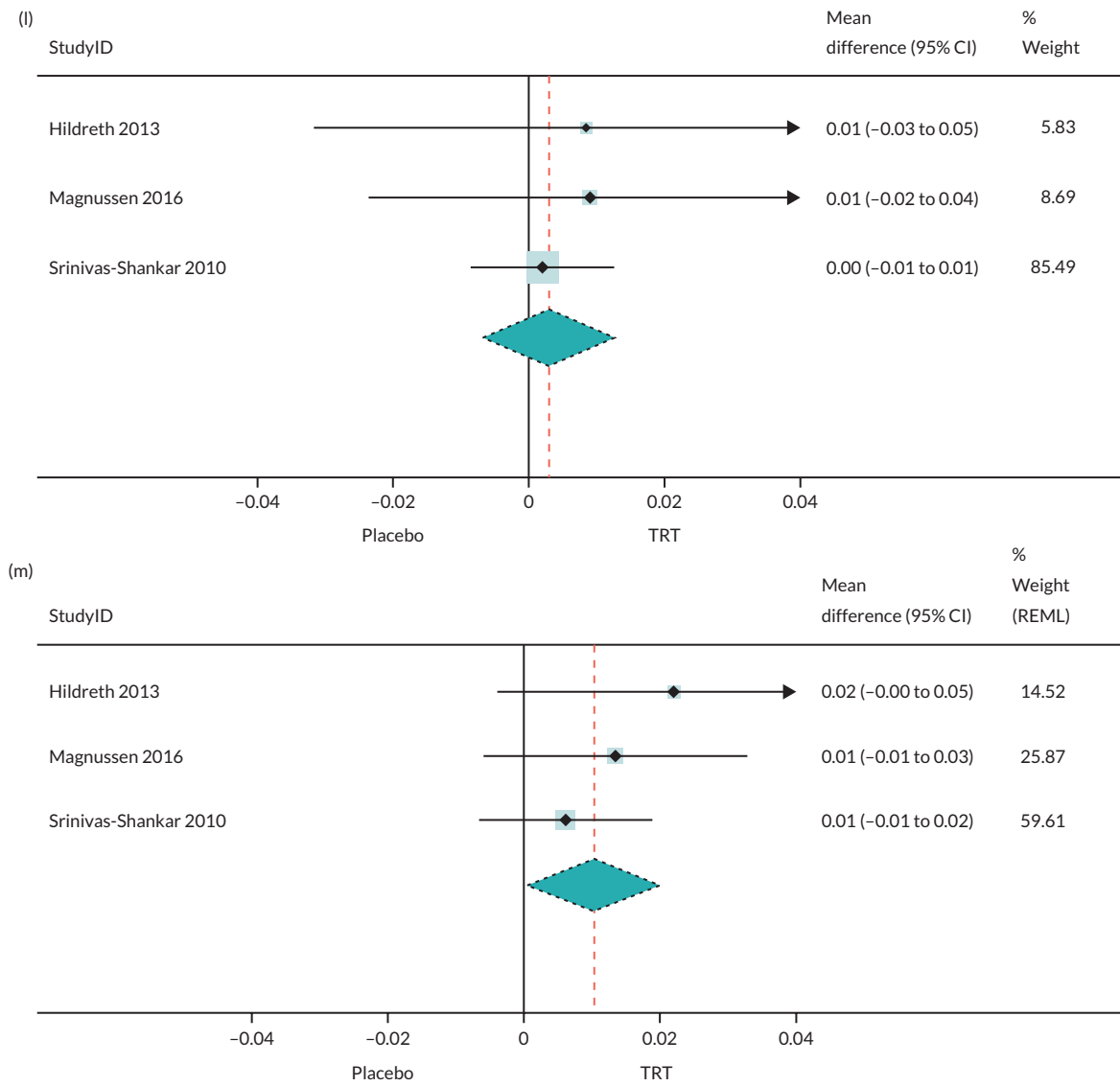


FIGURE 43 Two-stage meta-analysis for increase in bone mineral density (g/cm²). (a) Total score; (b) subtotal; (c) femoral neck; (d) lumbar spine; (e) thoracic spine; (f) total hip; (g) trochanter; (h) intertrochanter; (i) pelvis; (j) left arm; (k) right arm; (l) left leg; (m) right leg; (n) left rib; (o) right rib. a, Data presented as change from baseline. (continued)

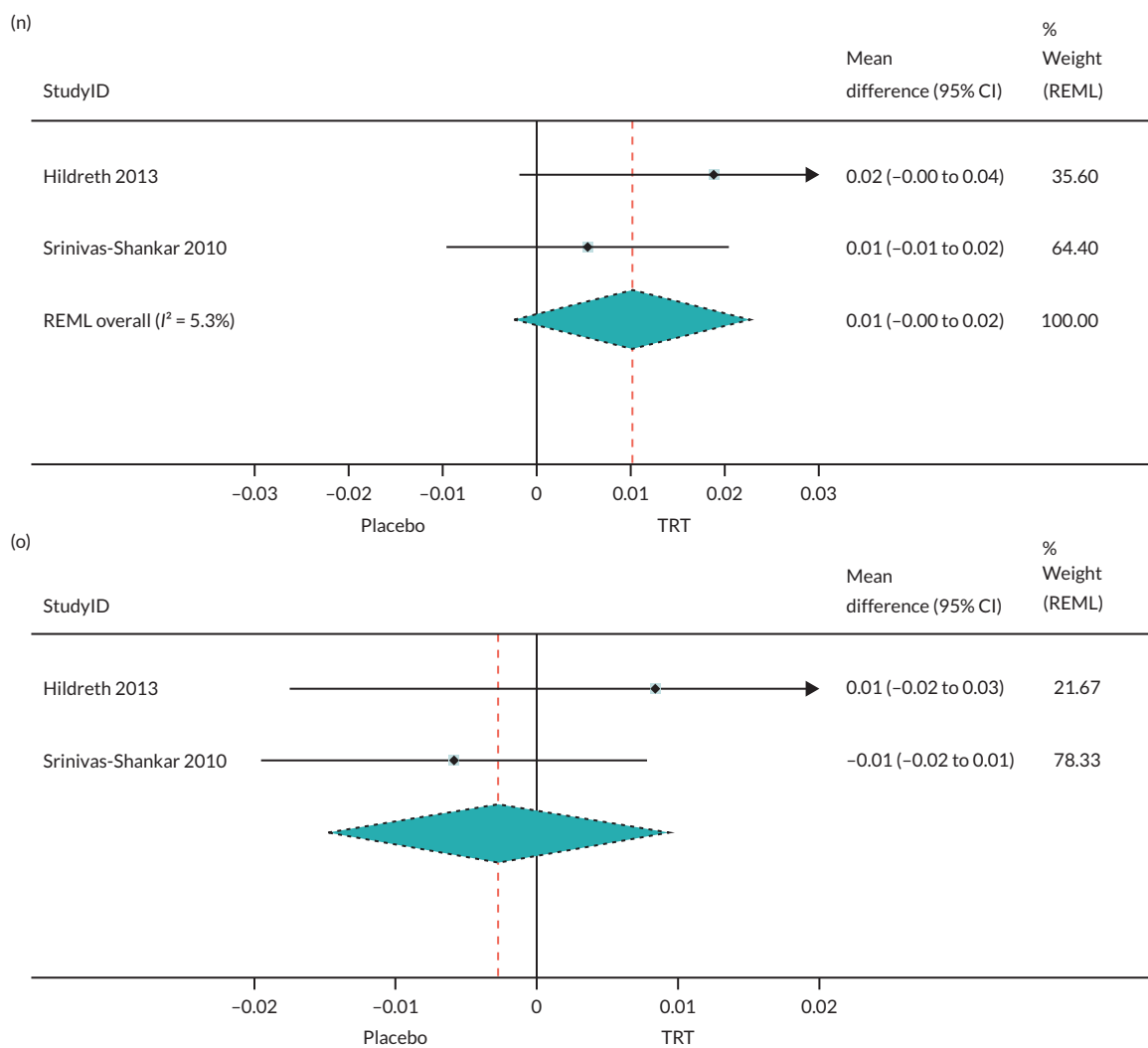


FIGURE 43 Two-stage meta-analysis for increase in bone mineral density (g/cm^2). (a) Total score; (b) subtotal; (c) femoral neck; (d) lumbar spine; (e) thoracic spine; (f) total hip; (g) trochanter; (h) intertrochanter; (i) pelvis; (j) left arm; (k) right arm; (l) left leg; (m) right leg; (n) left rib; (o) right rib. a, Data presented as change from baseline.

TABLE 45 One-stage meta-analysis analysis for psychological symptom outcomes

	Number of studies	TRT N = 1750	Placebo N = 1681	MD	95% CI	τ^2
BDI	3	6.99 (6.37); 143	8.49 (7.75); 103	-1.10	(-2.49 to 0.30)	0.71
Patient Health Questionnaire-9	1	4.04 (3.64); 336	4.55 (3.98); 319	-0.57	(-1.05 to -0.09)	
Positive and Negative Affect Scale						
Total	1	23.26 (3.85); 337	23.20 (3.66); 322	0.05	(-0.46 to 0.56)	
Positive	1	16.59 (3.73); 337	16.26 (3.39); 322	0.31	(-0.17 to 0.79)	
Negative	1	6.67 (2.25); 337	6.94 (2.60); 322	-0.25	(-0.54 to 0.05)	
Hospital Anxiety and Depression Scale (Depression)	1	6.76 (4.17); 80	6.73 (4.15); 96	-0.44	(-1.35 to 0.47)	
Centre for Epidemiologic Studies Depression Scale	1	5.81 (5.44); 47	7.92 (10.83); 24	-1.86	(-5.37 to 1.64)	

continued

Table 45 One-stage meta-analysis analysis for psychological symptom outcomes (continued)

	Number of studies	TRT N = 1750	Placebo N = 1681	MD	95% CI	τ^2
Aggression questionnaire						
Total	1	60.02 (12.37); 47	60.64 (16.71); 25	-0.54	(-5.20 to 4.11)	
Physical aggression	1	17.81 (6.06); 47	17.68 (5.65); 25	0.15	(-2.19 to 2.50)	
Verbal aggression	1	13.53 (4.32); 47	13.44 (3.27); 25	-0.48	(-2.06 to 1.10)	
Anger	1	14.09 (3.96); 47	14.36 (6.00); 25	-0.40	(-1.85 to 1.04)	
Hostility	1	14.60 (3.95); 47	15.16 (5.69); 25	0.21	(-1.38 to 1.80)	
Spielberger State-Trait Anxiety	1	23.47 (4.40); 19	30.89 (8.97); 19	-2.08	(-5.93 to 1.76)	

Notes

Values are mean (standard deviation); numbers.

Random-effects analysis was used for outcomes where more than one study provided IPD.

A fixed-effects analysis was used as only one study reported the outcome.

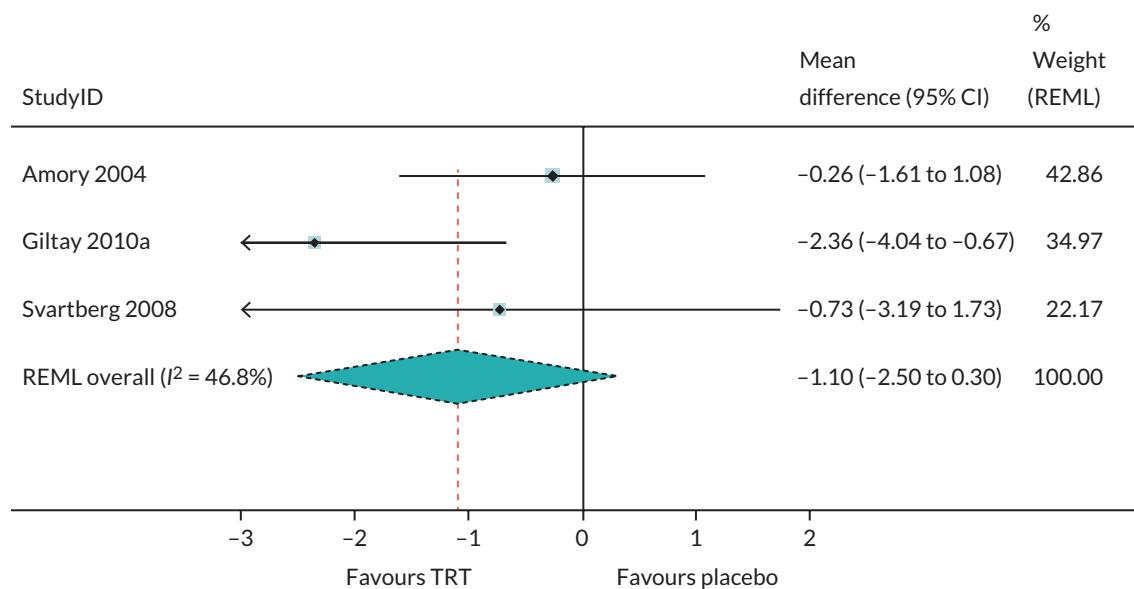


FIGURE 44 Two-stage meta-analysis for BDI.

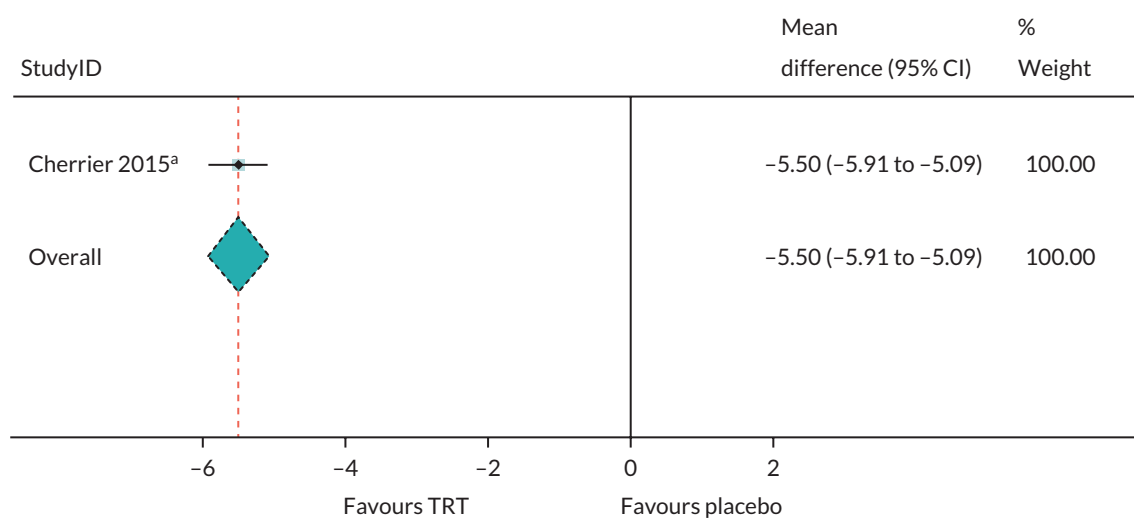


FIGURE 45 Two-stage meta-analysis for the Geriatric Depression Scale. a, Data presented as change from baseline.

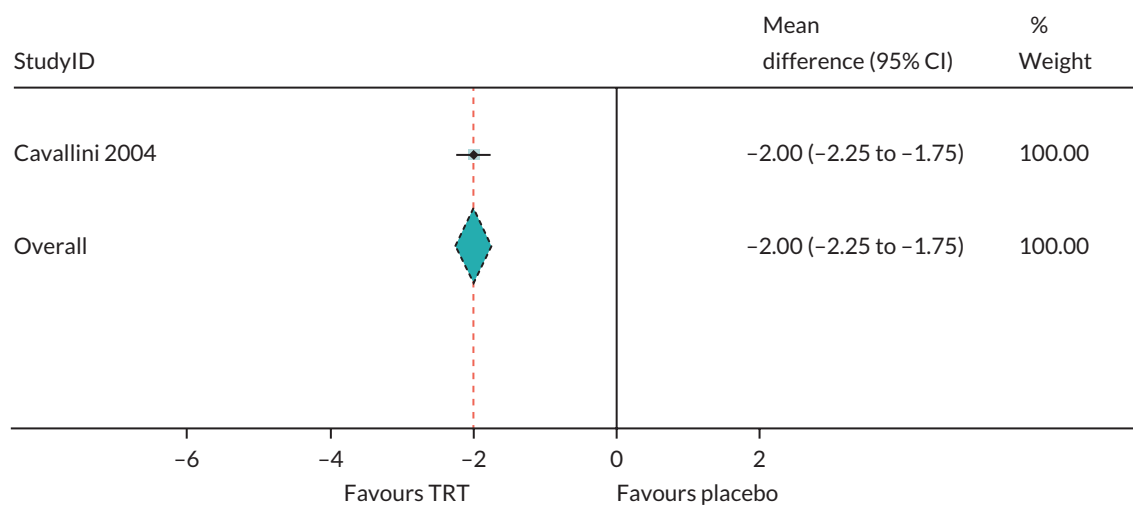


FIGURE 46 Two-stage meta-analysis for Hamilton Depression Scale.

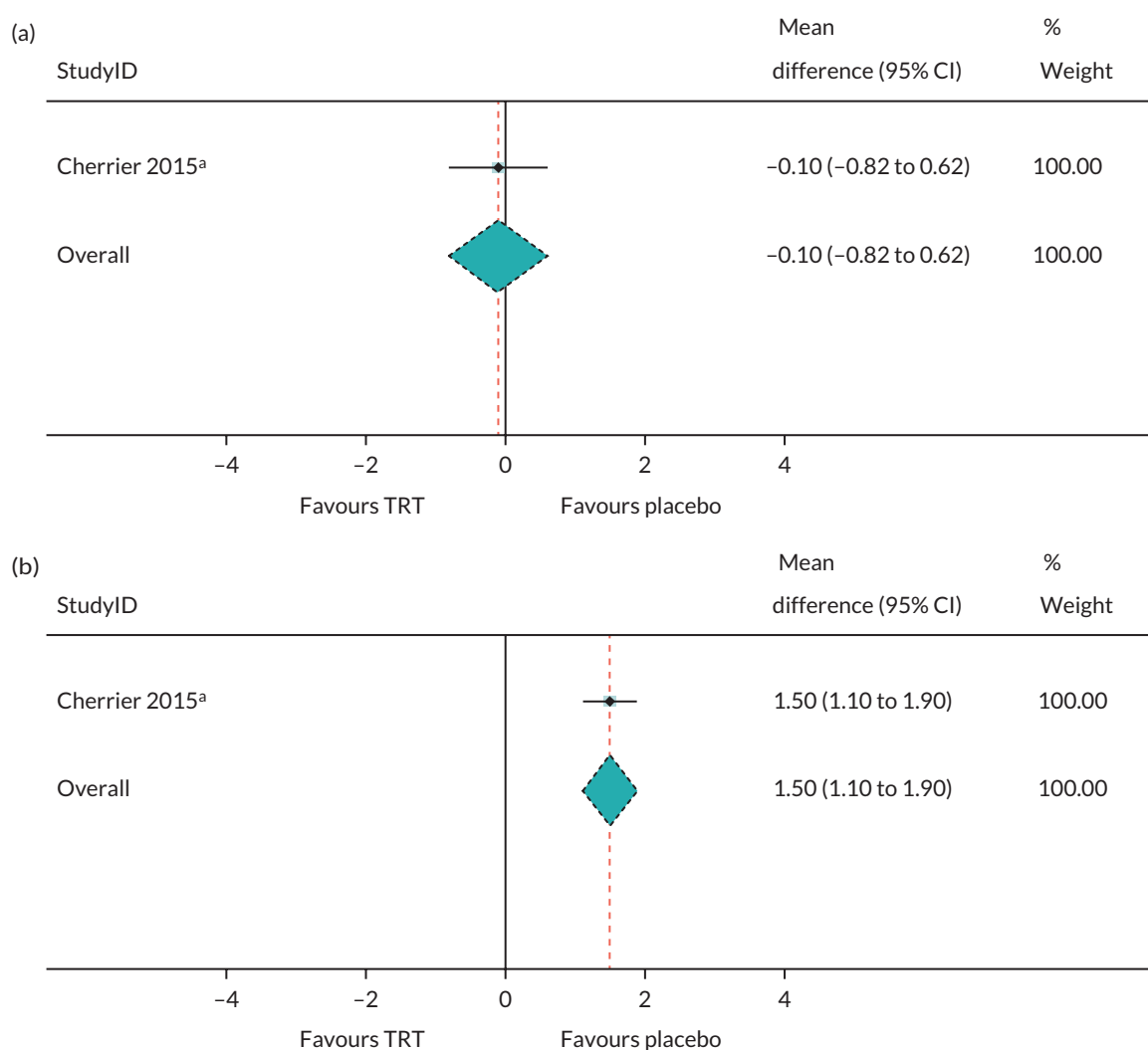


FIGURE 47 Two-stage meta-analysis for Profile of Mood States. (a) Anger; (b) depression; (c) fatigue; (d) tension and anxiety; (e) vigour. (continued)

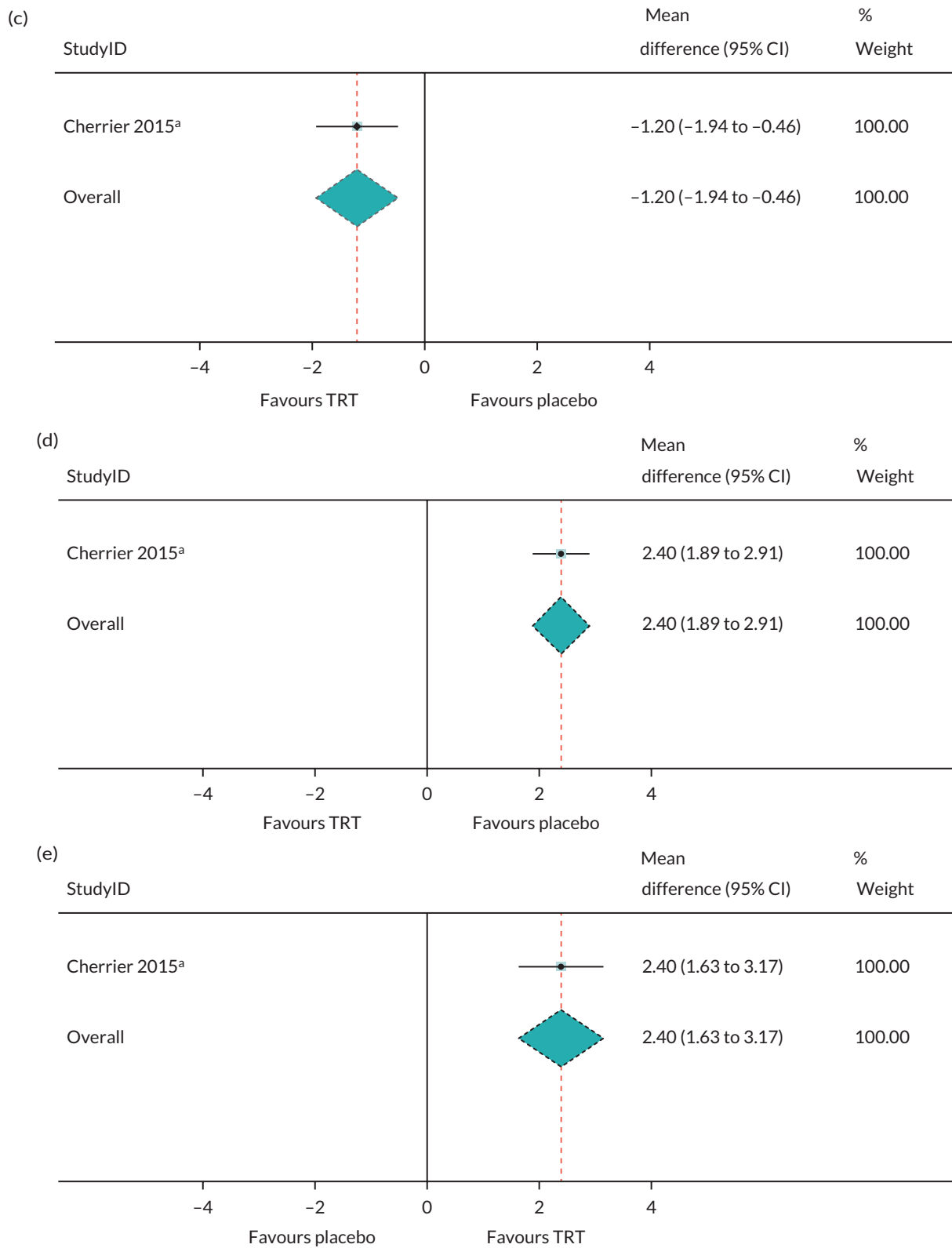


FIGURE 47 Two-stage meta-analysis for Profile of Mood States. (a) Anger; (b) depression; (c) fatigue; (d) tension and anxiety; (e) vigour. a, Data presented as change from baseline.

Appendix 7 Details of the health-related quality-of-life data and analysis for the economic model

SF-36 data from the TestES data set

[Table 46](#) shows the total number of participants in the data sets, which provided SF-36 data together with the collection time points. Data from Magnussen *et al.* 2016 for the 26-week assessment were grouped with the 26-week data from the other three studies and analysed together as one time point. A substantial number of participants presented missing data for all SF-36 questions, suggesting that the SF-36 data have been completed only for a subsample of the study (e.g. Basaria 2015).

Separate regression models were used to analyse all time point data. A mixed-effect regression model (random effects on participants and fixed effects on study) adjusted by baseline SF-6D score was used for the analysis of the 26-week data. Gamma regressions were used to obtain the difference in utility score at 52, 78 and 156 weeks and dummy variables were defined (1 for TRT and 0 for no TRT). [Table 47](#) shows non-statistically significant results for all time points. TRT dummy variable coefficient and standard error for the 26-week assessment were used to build a normal probability distribution used in the Markov model.

TABLE 46 SF-36 data available from received studies

Study	Number of participants in data set	Data collection time points and number of participants with any SF-36 data available at each time point					
		Baseline	24 weeks	26 weeks	52 weeks	78 weeks	156 weeks
Basaria 2015a	306	118		103		93	83
Emmelot-Vonk 2008	225	220		220			
Hildreth 2013	82	82		73	72		
Magnussen 2016	39	39	39				

TABLE 47 Utility score difference between TRT and no TRT for the SF-6D data analysis

Time point	N	Coefficient	Standard error	p-value	95% CI
26 weeks	409	0.0042	0.0084	0.4900	(-0.012 to 0.021)
52 weeks	72	0.0256	0.0242	0.2890	(-0.022 to 0.073)
78 weeks	91	-0.0355	0.0268	0.1840	(-0.088 to 0.017)
156 weeks	82	-0.0002	0.0229	0.9950	(-0.045 to 0.045)

Notes

BDI score data from the TestES data set.

Data sets from three studies reporting BDI data were available. The total number of participants in the data sets and the number of participants with BDI data at each time point are presented in [Table 44](#). Data were grouped for analysis into three groups according to the following time points: (1) 14- and 18-week, (2) 18-, 30- and 35-week and (3) 52-week time point.

TABLE 48 Beck Depression Inventory score (BDI) data available from received studies

Study	Number of participants in data set	Data collection time points and number of participants with BDI data available at each time point					
		14 weeks	18 weeks	28 weeks	30 weeks	35 weeks	52 weeks
Giltay	184		175		170		
Agledahl 2008	40			38			38
Amory 2004 ^a	70	44				39	38

a Amory 2004 also collected data at 4 weeks (45 participants) and 35 weeks (39 participants) 104 weeks (33 participants), and 156 weeks (34 participants).

Separate regression models were used to analyse all time point data. A mixed-effect regression model (random effects on participants and fixed effects on study) adjusted by baseline EQ-5D score was used for the analysis. Dummy variables were defined (1 for TRT and 0 for no TRT). [Table 45](#) shows non-statistically significant results for the 14- and 18-week data and for the 52-week time points. Statistically significant results were obtained at approximately 7 months (28–35 weeks). TRT dummy variable coefficient and standard error for this group were used to build a normal probability distribution used in the Markov model.

TABLE 49 Utility score difference between TRT and no TRT for the EQ0-5D score from the mapping exercise

Time point	N	Coefficient	Standard error	p-value	95% CI
14 and 18 weeks	219	0.0147	0.011	0.19	(-0.007 to 0.037)
28, 30 and 35 weeks	247	0.0295	0.009	0.00	(0.013 to 0.046)
52 weeks	76	0.0087	0.010	0.39	(-0.011 to 0.028)

EME
HSDR
HTA
PGfAR
PHR

Part of the NIHR Journals Library
www.journalslibrary.nihr.ac.uk

*This report presents independent research funded by the National Institute for Health and Care Research (NIHR).
The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the
Department of Health and Social Care*

Published by the NIHR Journals Library