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

Moira Cruickshank,[1] Jemma Hudson,[1] Rodolfo Hernández,[2] Magaly Aceves-Martins,[1] Richard Quinton,[3] Katie Gillies,[1] Lorna S Aucott,[1] Charlotte Kennedy,[2] Paul Manson,[1] Nicholas Oliver,[4] Frederick Wu,[5] Siladitya Bhattacharya,[6] Waljit S Dhillo,[4] Channa N Jayasena[4] and Miriam Brazzelli[1*]

[1]Health Services Research Unit, University of Aberdeen, Aberdeen, UK
[2]Health Economics Research Unit, University of Aberdeen, Aberdeen, UK
[3]Department of Endocrinology, Royal Victoria Infirmary, Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle, UK
[4]Faculty of Medicine, Imperial College London, London, UK
[5]Division of Diabetes, Endocrinology and Gastroenterology, University of Manchester, Manchester, UK
[6]School of Medicine, Medical Sciences and Nutrition, University of Aberdeen, Aberdeen, UK

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

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

Scientific summary

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

Health Technology Assessment 2024; Vol. 28: No. 43
DOI: 10.3310/JRYT3981

NIHR Journals Library www.journalslibrary.nihr.ac.uk

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.

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.

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