Oral testosterone replacement in symptomatic late-onset hypogonadism: effects on rating scales and general safety in a randomized, placebo-controlled study

Jean-Jacques Legros, Eric J H Meuleman, Jolanda M H Elbers, T B Paul Geurts, Marion J G H Kaspers, Pierre M G Bouloux for the Study 43203 Investigators

Department of Endocrinology, University of Liège, CHR de la Citadelle, Boulevard du 12e de Ligne 1, 4000 Liège, Belgium, 1Department of Urology, Free University Medical Center, PO Box 7057, 1007 MB Amsterdam, The Netherlands, 2Global Clinical Research, 3Global Communications R&D and 4Research Data and Quantitative Sciences, Schering-Plough, PO Box 20, 5340 BH, Oss, The Netherlands and 5Centre for Neuroendocrinology, Royal Free and University College School of Medicine, 12th Floor Cockayne Ward, Pond Street, London NW3 2QG, UK

(Correspondence should be addressed to J-J Legros; Email: jean-jacques.legros@ulg.ac.be)

Abstract

Objective: To investigate the effects of oral testosterone undecanoate (TU) on symptoms associated with late-onset hypogonadism (LOH).

Design: Multicenter, randomized, double-blind, placebo-controlled.

Methods: The study was performed in 14 study centers in seven European countries. Men ≥50 years (n = 322) with symptoms of hypogonadism and testosterone deficiency (calculated free testosterone <0.26 nmol/l) were randomized and treated for 12 months with placebo or oral TU 80, 160 or 240 mg/day. Primary outcome was the total score on the Aging Males’ Symptoms (AMS) rating scale after six months of treatment.

Results: Treatment of mild-to-moderate LOH symptoms in subjects with borderline hypogonadism with oral TU resulted in an improved total AMS score at month 6, but differences between groups were not statistically significant. There was greater improvement in subjects >60 years when compared with subjects ≤60 years (P = 0.001), but baseline testosterone level had no influence on treatment response. The AMS sexual symptoms domain improved with oral TU 160 mg/day at months 6 (P = 0.008) and 12 (P = 0.012) compared with placebo, but not with 80 and 240 mg/day. Treatment was well-tolerated and there were no between-group differences in adverse events or drop-out rates.

Conclusions: In one of the largest placebo-controlled studies of testosterone therapy in LOH, oral TU did not improve total AMS score in subjects with mild-to-moderate symptoms compared with placebo, except the sexual symptom sub-domain where a modest improvement was reported with oral TU 160 mg/day.

European Journal of Endocrinology 160 821–831

Introduction

Late-onset hypogonadism (LOH) and its treatment have become a topic of increasing interest in endocrinology and urology, particularly in the wake of the availability of a number of new testosterone formulations (patches, gels, buccal, and long-acting injectables) since 1995 and the introduction of new treatment guidelines issued by scientific associations (1, 2). These guidelines acknowledge that LOH is a true clinical entity that can result in bone loss, reduced lean body mass and strength, sexual dysfunction, mood swings, low energy, increased body fat, anemia, and impaired quality of life. The underlying cause of LOH can often be attributed to a combination of testicular failure and hypothalamic–pituitary dysfunction. Guidelines also agree that only those men should be treated who present with sub-normal serum testosterone levels in combination with clinical signs and symptoms of hypogonadism (1). Recently, the prospective population-based Rancho Bernardo Study reported that the low testosterone level in older men was associated with increased mortality (3). A number of randomized controlled trials have also demonstrated the clinical efficacy of various testosterone preparations in management of signs and symptoms associated with LOH (4–14). However, most of these studies are handicapped by short duration and/or small size, lack of uniform definition of LOH, and in most trials dose–response relationships have not been investigated.

Oral testosterone undecanoate (TU) capsules have been available since the end of the 1970s (Andriol) and were recently reformulated (Andriol Testocaps) to improve pharmaceutical stability, allowing for more
practical storage conditions. It is the only preparation for oral use enabling delivery of testosterone to the systemic circulation via lymphatic absorption (15, 16). Several pharmacokinetic studies have shown that the new formulation of oral TU, when taken with a normal meal, generates a dose-dependent increase in serum testosterone levels to levels within the physiological range (17–19). Although evidence is relatively scarce, it has been proposed that oral TU might in particular be suited for elderly patients because the oral administration route is convenient, dose can easily be titrated to individual needs and the preparation can be discontinued quickly as needed (1). The objective of this study was therefore to investigate the effects of different doses of oral TU on clinical outcomes in symptomatic aging men with androgen deficiency.

**Subjects and methods**

The study was a multicenter, randomized, double-blind, placebo-controlled clinical trial, involving 14 centers in Europe, located in Austria (1 site), Belgium (3 sites), France (1 site), Germany (2 sites), UK (3 sites), The Netherlands (3 sites), and Switzerland (1 site). The trial was conducted between November 2001 (first subject entered) and July 2004 (last assessment of the last subject).

**Clinical trial registry**

This trial was registered at www.clinicaltrials.gov under number NTC00434824.

**Subjects**

Subjects were men at least 50 years of age with a body mass index (BMI) between 18 and 34 kg/m², with symptoms of androgen deficiency (as indicated by a positive score on androgen deficiency in aging males (ADAM) questionnaire) (20) as well as a calculated free morning testosterone measurement of <0.26 nmol/l (21). This cut-off level was determined on the basis of 2 s.d.s below the normal range in young men, as assessed in a sample of 150 men (22).

The most important reasons to exclude potential study participants prior to randomization included: history or present diagnosis of breast or prostate cancer; any clinically significant abnormal finding on physical examination including the prostate; lower urinary tract symptoms (LUTS; international prostate symptom score (IPSS) > 14); a prostate specific antigen (PSA) level > 4 ng/ml at screening; cause of androgen deficiency other than aging; treatment with androgens or anabolic steroids within the last two years; hyperprolactinemia or treatment with prolactin-lowering drugs; history of known chronic polycythemia and/or hematocrit > 50% at screening; history or presence of severe sleep apnea; uncontrolled hypertension; insulin-dependent diabetes mellitus; any clinically significant abnormal hematological or biochemical values or physical finding; unstable or untreated endocrine disorders; liver or renal diseases; psychiatric diseases; history or presence of other disorders such as clinically significant cardiovascular, cerebrovascular or hematological disorders which in the opinion of the investigator, could compromise the subject’s participation in the trial. Subjects who used medication that would interfere with the efficacy and safety objectives of the trial or who were not stabilized on any concomitant medication (continuous use) during at least three months prior to randomization, could not participate. Study participants were recruited using media (advertisements, radio shows, etc.) or were recruited from patient pools at male health clinics, urology departments, etc.

**Intervention**

In this trial, subjects were treated for 12 months with oral TU capsules, each containing 40 mg TU (Andriol Testocaps) or identical placebo. Subjects were randomized to one of the four treatment groups:

1. Two placebo capsules in the morning, two placebo capsules in the afternoon, and two placebo capsules in the evening, immediately after meals;
2. Two capsules of TU in the morning, two placebo capsules in the afternoon, and two placebo capsules in the evening, immediately after meals (total daily dose: 80 mg TU);
3. Two capsules of TU in the morning, two placebo capsules in the afternoon, and two capsules of TU in the evening, immediately after meals (total daily dose: 160 mg TU);
4. Two capsules of TU in the morning, two capsules of TU in the afternoon, and two capsules of TU in the evening, immediately after meals (total daily dose: 240 mg TU).

Subjects were assigned a subject code number in keeping with the order of their enrollment in the clinical trial, i.e. the first subject received the first (lowest) number and each subsequent subject received an increasing number, with a block size of four.

**Methods**

The study was executed in accordance to the Declaration of Helsinki, ICH Guideline for Good Clinical Practice, and applicable regulatory guidelines. During screening, subjects were informed, orally and in writing, about the purposes, procedures, and general risks of participating in the trial. Demographics were recorded and a general physical examination was performed, a medical history was taken and blood was drawn for laboratory analyses. In order to investigate
eligibility for inclusion in the trial, subjects were asked to complete the ADAM questionnaire. A positive questionnaire was defined as a ‘yes’ to questions 1 or 7 or any three other questions (20). Furthermore, blood samples were collected in the morning (between 0600 and 1000 h) to determine levels of total testosterone and sex hormone-binding globulin (SHBG) for calculation of free testosterone using the formula by Vermeulen et al. (21). The screening assessments of total testosterone and SHBG were done by ABL (Assen, The Netherlands) using DELFIA, time-resolved fluorometric immunoassay (Wallac Oy, Turku, Finland). Analysis of endocrine parameters in samples taken at baseline and during the in-treatment period was performed by the Bioanalytics Department at Essex Pharma Development GmbH (Waltrop, Germany). Serum testosterone was measured with a validated gas chromatography with mass spectrometric detection, whereas SHBG was measured using a Perkin Elmer AutoDELFIA fluor-immunometric assay.

Questionnaires

Eligible subjects were asked to complete the following health-related quality of life (HRQoL) questionnaires: the AMS rating scale (23) (Baseline, months 1, 3, 6, 9, and 12), Derogatis Interview for Sexual Function – Self-Report (DISF-SR) (24), and the Medical Outcomes Study Short Form-36 (SF-36; Baseline, months 3, 6, and 12) (25). The change from baseline on the total score of the AMS rating scale at month 6 was predefined as the primary efficacy parameter. The AMS rating scale consists of 17 symptoms evaluated on a 5-point scale (1 = none, 5 = extremely severe). The 17 symptoms are distributed over three sub-domains: psychological (five symptoms), somatic (seven symptoms), and sexual (five symptoms). Symptom severity of the total AMS score is defined as ‘no/little’ (17–26 points), ‘mild’ (27–36 points), ‘moderate’ (37–49 points), and ‘severe’ (≥ 50 points).

Safety assessments

A general physical examination, including testicular and prostate examination (digital rectal examination or ultrasound), was performed at screening, month 6 (prostate only) and month 12. The IPSS was assessed at baseline, month 6 and month 12. Heart rate, blood pressure, and body weight were measured at screening, months 1, 3, 6, 9, and 12. A full panel of routine laboratory parameters were done at screening and month 12, with a limited panel (hemoglobin, hematocrit, erythrocyte count, alanine aminotransferase (ALAT), aspartate aminotransferase (ASAT), bilirubin, creatinine, gamma glutamyl transferase (GGT), PSA) also at months 1, 3, 6, and 9. Information regarding adverse events (AEs) was obtained by questioning or examining the subjects, and included all new complaints and symptoms and pre-existing complaints or symptoms that increased in severity or frequency during the treatment period.

Statistical analysis

A two-way ANOVA was used to compare the change from baseline in each dose group with placebo. This method was used for the AMS total scores and its sub-domains, DISF-SR and SF-36. An outcome was considered statistically significant if P ≤ 0.05. Exploratory subgroup analyses were performed for the AMS parameters stratified by age (<60 and ≥ 60 years), by baseline total testosterone concentrations (≤ 10.4 vs > 10.4 nmol/l) and by symptom severity at baseline. Descriptive statistics were calculated for each efficacy parameter by group and assessment.

With respect to the safety analysis, descriptive statistics were calculated for AEs, laboratory parameters, physical examination (including testicular and prostate examination), vital signs, and some special safety considerations (PSA > 4 ng/ml, hematocrit > 54%, IPSS > 14). Additionally, AEs occurring in more than 4% in any of the groups were tested using Fisher’s exact tests.

Results

Disposition of subjects

A total of 1444 subjects were screened and of these, 322 were randomized. The most common screening failure was having a calculated free testosterone level above the cut-off level of 0.26 nmol/l, despite a positive ADAM score. Of the 322 randomized subjects, six never received treatment. As a consequence, 316 subjects received at least one dose of trial medication (all-subjects-treated group) and were included in the safety analyses. Efficacy analyses were performed on the 310 subjects who had at least one post-baseline assessment with respect to the AMS total score (intent-to-treat group). Of all randomized subjects, 243 completed the trial (completion rate: 75%), without marked differences between groups with respect to completion rate or reasons for premature discontinuation. Most common reasons for discontinuation were occurrence of AEs (n = 39) or withdrawal of consent (n = 22; Fig. 1). Overall, the mean age of the subjects was 58.7 years with a BMI of 27.3 kg/m². Almost all subjects (98.7%) were Caucasians and the majority of the subjects were married at the time they entered the trial. Approximately less than half of the subjects were employed full-time at screening and about one-third were retired. There were no relevant differences between the groups with respect to age, body weight, height, BMI, education, marital status or racial composition (Table 1).
The randomized subjects had at screening a mean total testosterone level of 12.8 ± 4.2 nmol/l (range: 3.5–35.7 nmol/l) and a mean SHBG of 45.6 ± 19.8 nmol/l (range: 12.5–164.0 nmol/l), resulting in a mean calculated free testosterone of 0.21 ± 0.04 nmol/l (range: 0.05–0.26 nmol/l). At baseline, the mean total testosterone level was 14.9 ± 5.5 nmol/l (range: 4.1–37.4 nmol/l) and mean SHBG was 44.0 ± 19.2 nmol/l (range: 14.2–138.0 nmol/l). There were no relevant differences between the treatment groups at screening or at baseline. Dosing compliance was high, with a median of 98.7% of all capsules taken according to the study protocol. There were no differences in compliance between the four treatment groups or between morning, afternoon or evening intake.

**Symptoms**

At baseline, the mean score on the AMS rating scale ranged from 37.3 to 39.8 (Table 2). After six months of treatment (at the time of primary efficacy assessment), an improvement of symptoms was seen in all groups (on average −4.2 with placebo, −4.8 with oral TU 80 mg/day, −5.6 with oral TU 160 mg/day and −5.4

**Table 1** Comparison of the demographic data at screening (all-subjects-treated group).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Placebo n=79</th>
<th>80 mg/day n=78</th>
<th>160 mg/day n=82</th>
<th>240 mg/day n=77</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years) Mean±s.d.</td>
<td>58.4 ± 5.5</td>
<td>59.5 ± 6.5</td>
<td>58.4 ± 5.7</td>
<td>58.6 ± 5.7</td>
</tr>
<tr>
<td>Body weight (kg) Mean±s.d.</td>
<td>84.9 ± 10.3</td>
<td>85.5 ± 12.9</td>
<td>87.1 ± 12.7</td>
<td>85.1 ± 13.6</td>
</tr>
<tr>
<td>Body height (cm) Mean±s.d.</td>
<td>176.0 ± 6.5</td>
<td>176.8 ± 6.3</td>
<td>177.8 ± 6.5</td>
<td>177.9 ± 7.4</td>
</tr>
<tr>
<td>Body mass index (kg/m²) Mean±s.d.</td>
<td>27.4 ± 3.0</td>
<td>27.3 ± 3.4</td>
<td>27.5 ± 3.5</td>
<td>26.8 ± 3.5</td>
</tr>
<tr>
<td>Marital status (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never married</td>
<td>4 (5.1%)</td>
<td>5 (6.4%)</td>
<td>3 (3.7%)</td>
<td>3 (3.9%)</td>
</tr>
<tr>
<td>Presently married</td>
<td>66 (83.5%)</td>
<td>61 (78.2%)</td>
<td>67 (81.7%)</td>
<td>64 (83.1%)</td>
</tr>
<tr>
<td>Living together</td>
<td>4 (5.1%)</td>
<td>4 (5.1%)</td>
<td>5 (6.1%)</td>
<td>4 (5.2%)</td>
</tr>
<tr>
<td>Divorced/separated</td>
<td>4 (5.1%)</td>
<td>7 (9.0%)</td>
<td>5 (6.1%)</td>
<td>5 (6.5%)</td>
</tr>
<tr>
<td>Widowed</td>
<td>1 (1.3%)</td>
<td>1 (1.3%)</td>
<td>2 (2.4%)</td>
<td>1 (1.3%)</td>
</tr>
<tr>
<td>Education (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary school</td>
<td>2 (2.5%)</td>
<td>3 (3.8%)</td>
<td>5 (6.1%)</td>
<td>6 (7.8%)</td>
</tr>
<tr>
<td>High school</td>
<td>20 (25.3)</td>
<td>27 (34.6%)</td>
<td>25 (30.5%)</td>
<td>22 (28.6%)</td>
</tr>
<tr>
<td>Apprenticeship</td>
<td>32 (40.5%)</td>
<td>26 (33.3%)</td>
<td>24 (29.3%)</td>
<td>28 (36.4%)</td>
</tr>
<tr>
<td>College</td>
<td>25 (31.6%)</td>
<td>22 (28.2%)</td>
<td>28 (34.1%)</td>
<td>21 (27.3%)</td>
</tr>
<tr>
<td>Race (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>2 (2.5%)</td>
<td>0 (0.0%)</td>
<td>0 (0.0%)</td>
<td>0 (0.0%)</td>
</tr>
<tr>
<td>Caucasian</td>
<td>75 (94.9%)</td>
<td>78 (100.0%)</td>
<td>82 (100.0%)</td>
<td>77 (100.0%)</td>
</tr>
<tr>
<td>Other</td>
<td>2 (2.5%)</td>
<td>0 (0.0%)</td>
<td>0 (0.0%)</td>
<td>0 (0.0%)</td>
</tr>
</tbody>
</table>
with oral TU 240 mg/day; all statistically significant as compared with baseline), but the between-group differences were not statistically significant (Fig. 2). With continued treatment up to 12 months, scores on the AMS rating scale remained improved as compared with baseline in all four groups (P < 0.05), but the improvement in the active groups was not superior compared with placebo. Restricting the analysis to subjects with moderate-to-severe symptoms at baseline did not change the results. After six months of treatment, there was a stronger improvement of the AMS total score in subjects < 60 years as compared with subjects ≥ 60 years (P = 0.001). There were no differences in mean AMS total scores after six months of treatment between subjects with baseline testosterone ≤ 10.4 nmol/l as compared with subjects with baseline testosterone > 10.4 nmol/l.

Also the scores in the AMS sub-domains ‘psychological symptoms’, ‘somatic symptoms’, and ‘sexual symptoms’ improved in all four treatment groups, which were statistically significant for sexual symptoms at month 6 with oral TU 160 mg/day (P = 0.008; Table 2) as well as for the combined oral TU groups (P = 0.044) as compared with placebo. Improvements of AMS subscales were generally stronger in the active groups than with placebo, but there were no statistically significant between-group differences. With continued treatment, there was also an improvement of the AMS sexual symptom sub-domain with oral TU 160 mg/day as compared with placebo at months 9 (P = 0.022) and 12 (P = 0.012; Fig. 3). Treatment with 80 mg/day or 240 mg/day of oral TU did not improve the AMS sexual symptoms sub-scale compared with placebo.

**Sexual function and HRQoL**

Sexual function total score on the DISF-SR scale improved in all four groups. The improvements with oral TU on the (sub-) scores were generally not different

---

**Table 2** Effects of oral testosterone undecanoate (TU) and placebo on the Aging Males’ Symptoms (AMS) rating scale (plus sub-domains) after 6 and 12 months of treatment (mean ± s.e.m.; intent-to-treat group).

<table>
<thead>
<tr>
<th>Sub-score</th>
<th>Placebo (n = 79)</th>
<th>80 mg/day (n = 78)</th>
<th>160 mg/day (n = 82)</th>
<th>240 mg/day (n = 77)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AMS rating scale – total score</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>39.8 ± 10.1</td>
<td>38.0 ± 11.4</td>
<td>37.3 ± 10.1</td>
<td>37.6 ± 11.6</td>
</tr>
<tr>
<td>Month 6</td>
<td>35.6 ± 9.5</td>
<td>33.2 ± 10.2</td>
<td>31.7 ± 9.9</td>
<td>32.2 ± 9.5</td>
</tr>
<tr>
<td>∆ from baseline</td>
<td>−4.2 ± 8.5</td>
<td>−4.8 ± 8.0</td>
<td>−5.6 ± 9.9</td>
<td>−5.4 ± 10.3</td>
</tr>
<tr>
<td>Month 12</td>
<td>35.6 ± 9.7</td>
<td>33.3 ± 10.2</td>
<td>31.7 ± 8.9</td>
<td>32.6 ± 10.3</td>
</tr>
<tr>
<td>∆ from baseline</td>
<td>−4.2 ± 9.1</td>
<td>−4.7 ± 8.1</td>
<td>−5.6 ± 8.8</td>
<td>−5.0 ± 10.4</td>
</tr>
<tr>
<td><strong>AMS rating scale – psychological symptoms sub-score</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>10.8 ± 3.8</td>
<td>10.1 ± 4.2</td>
<td>9.6 ± 3.9</td>
<td>9.7 ± 4.3</td>
</tr>
<tr>
<td>Month 6</td>
<td>9.6 ± 3.5</td>
<td>8.8 ± 3.8</td>
<td>8.4 ± 3.6</td>
<td>8.3 ± 3.5</td>
</tr>
<tr>
<td>∆ from baseline</td>
<td>−1.2 ± 3.3</td>
<td>−1.2 ± 3.0</td>
<td>−1.2 ± 3.1</td>
<td>−1.3 ± 3.2</td>
</tr>
<tr>
<td>Month 12</td>
<td>9.3 ± 3.7</td>
<td>8.5 ± 3.5</td>
<td>8.3 ± 3.2</td>
<td>8.3 ± 3.6</td>
</tr>
<tr>
<td>∆ from baseline</td>
<td>−1.4 ± 3.6</td>
<td>−1.5 ± 2.9</td>
<td>−1.3 ± 3.1</td>
<td>−1.4 ± 3.4</td>
</tr>
<tr>
<td><strong>AMS rating scale – somatic symptoms sub-score</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>15.6 ± 4.6</td>
<td>15.2 ± 5.1</td>
<td>14.3 ± 4.4</td>
<td>15.0 ± 5.4</td>
</tr>
<tr>
<td>Month 6</td>
<td>13.6 ± 4.8</td>
<td>13.1 ± 4.3</td>
<td>12.6 ± 4.4</td>
<td>12.8 ± 4.2</td>
</tr>
<tr>
<td>∆ from baseline</td>
<td>−2.0 ± 4.0</td>
<td>−2.1 ± 3.7</td>
<td>−1.7 ± 4.4</td>
<td>−2.3 ± 4.8</td>
</tr>
<tr>
<td>Month 12</td>
<td>13.8 ± 4.7</td>
<td>13.3 ± 4.7</td>
<td>12.4 ± 3.8</td>
<td>13.0 ± 4.5</td>
</tr>
<tr>
<td>∆ from baseline</td>
<td>−1.8 ± 4.1</td>
<td>−2.0 ± 4.0</td>
<td>−1.9 ± 4.0</td>
<td>−2.0 ± 4.7</td>
</tr>
<tr>
<td><strong>AMS rating scale – sexual symptoms sub-score</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>13.5 ± 3.7</td>
<td>12.7 ± 3.7</td>
<td>13.3 ± 3.8</td>
<td>12.9 ± 3.9</td>
</tr>
<tr>
<td>Month 6</td>
<td>12.4 ± 3.9</td>
<td>11.3 ± 4.1</td>
<td>10.7 ± 4.0</td>
<td>11.0 ± 4.2</td>
</tr>
<tr>
<td>∆ from baseline</td>
<td>−1.1 ± 3.7</td>
<td>−1.4 ± 3.0</td>
<td>−2.6 ± 4.1</td>
<td>−1.9 ± 3.8</td>
</tr>
<tr>
<td>Month 12</td>
<td>12.5 ± 3.9</td>
<td>11.5 ± 3.9</td>
<td>11.0 ± 3.9</td>
<td>11.1 ± 4.2</td>
</tr>
<tr>
<td>∆ from baseline</td>
<td>−0.9 ± 3.4</td>
<td>−1.2 ± 3.0</td>
<td>−2.3 ± 3.9</td>
<td>−1.8 ± 3.7</td>
</tr>
</tbody>
</table>

*P < 0.05; †P < 0.01 versus change in placebo.
from the change in placebo, except for improvements on the total DISF-SR score ($P < 0.044$), sexual behavior/experience ($P < 0.029$), and orgasms ($P < 0.031$) with oral TU 240 mg/day at month 3. There were no statistically significant improvements in any of the domains of HRQoL as measured by the SF-36 in any of the treatment groups (results not shown).

**Safety evaluations**

A total of 254 out of 316 subjects (80.4%) reported one or more AEs. There were no clinically relevant differences in overall AE rates between groups. Two deaths were reported: one subject in the oral TU 160 mg/day group died during treatment (Day 148) due to arrhythmia and hypertrophic cardiomyopathy (not related to the study drug as judged by the investigator). Another subject in the oral TU 240 mg/day group died 3 months after the end of treatment (Day 453) due to cardiac arrest (unlikely related to the intervention as judged by the investigator). As shown in Table 3, two subjects (2.5%) in the placebo group, seven subjects (9.0%) in the oral TU 80 mg/day group, seven subjects (8.5%) in the oral TU 160 mg/day, and three subjects (3.9%) in the oral TU 240 mg/day group experienced at least one serious AE. Most of these were (according to the investigator) not/unlikely related to the study drug, with the exception of one case of prostate cancer which was detected after 10 months (Day 326) in the oral TU 80 mg/day group which was rated as possibly related to the intervention. AEs that occurred with a prevalence of > 5% in any of the treatment groups were: influenza, nasopharyngitis, headache, conjunctivitis, hypertension, cough, abdominal pain, diarrhea, dyspepsia, arthralgia, back pain, localized osteoarthritis, myalgia, erectile dysfunction, fatigue, PSA increase, and weight increase. Conjunctivitis was reported in 5.1% of subjects with placebo and not with oral TU ($P = 0.011$); back pain occurred in 1.3% with placebo, in 11.5% with oral TU 80 mg/day, in 6.1% with oral TU 160 mg/day and in 13.0% with oral TU 240 mg/day ($P = 0.014$); localized osteoarthritis was reported in 1.3% with placebo and in 5.1% with oral TU 80 mg/day ($P = 0.025$). The percentage of subjects reporting diarrhea was between 5.1 and 7.7% with no differences between the treatment groups.

During the study, there were minor changes in the mean serum PSA level during treatment in all groups. The mean percentage change from baseline at month 12 was highest for the placebo group (i.e., 17.4%) and varied from 4.2% to 12.9% in the active treatment groups. There was no difference between groups on the incidence of clinically relevant PSA elevations (Table 4). Moreover, no adverse changes were observed on the IPSS and there were no differences between groups in the incidence of clinically significant prostate symptoms, defined as an IPSS $> 14$ (Table 4). There was a dose-dependent increase of hematocrit with oral TU, indicated by a mean percentage change from baseline at months 12 of 1.0% for the 80 mg/day group, 2.1% in the 160 mg/day group and 5.2% in the 240 mg/day group compared with $-0.5$% in the placebo group. The incidence of a clinically relevant elevation of hematocrit was higher with oral TU 240 mg/day (6.6%) as compared with the other active treatment groups (2.6% and 1.3% in the 80 mg/day group and 160 mg/day group respectively), while none of the subjects in the placebo group had a hematocrit level

![Figure 3](assets/fig3.png)

**Figure 3** Effect of oral TU on the AMS rating scale – sexual symptoms score (mean change from baseline in the intent-to-treat group).

### Table 3 Overview of subjects with at least one adverse event (all-subjects-treated group).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Placebo n=79</th>
<th>Oral testosterone undecanoate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>80 mg/day n=78</td>
</tr>
<tr>
<td>Subjects with AEs</td>
<td>65, 82.3%</td>
<td>70, 76.9%</td>
</tr>
<tr>
<td>Deaths</td>
<td>0, 0.0%</td>
<td>0, 0.0%</td>
</tr>
<tr>
<td>Subjects with serious AEs (SAEs)</td>
<td>2, 2.5%</td>
<td>7, 9.0%</td>
</tr>
<tr>
<td>Discontinuations due to AEs</td>
<td>10, 12.7%</td>
<td>7, 9.0%</td>
</tr>
<tr>
<td>Subjects with drug-related AEs</td>
<td>27, 34.2%</td>
<td>31, 39.7%</td>
</tr>
<tr>
<td>Subjects with AEs of known severe intensity</td>
<td>13, 16.5%</td>
<td>10, 12.8%</td>
</tr>
</tbody>
</table>

www.eje-online.org
above 54% (Table 4). There were no adverse effects of oral TU on liver function as assessed by measurement of serum levels of ALAT, ASAT, bilirubin, and GGT.

**Discussion**

In this trial – one of the largest studies so far in symptomatic aging men with androgen deficiency – the effect of oral TU on the total AMS score during 12 months of treatment was not significantly different from placebo, except for sexual symptoms where a modest improvement was reported with oral TU 160 mg/day. Treatment with any of the investigated doses was generally well-tolerated and safe.

**Symptoms**

The primary outcome of this study was the change on the total score of the AMS rating scale after six months of treatment. The absence of a statistically significant effect of oral TU on total AMS score compared with placebo at month 6 in this study may be due to the following reasons: a) the AMS rating scale is not a useful tool for assessing efficacy of testosterone in LOH, b) the criteria used for diagnosis of LOH are inaccurate, c) testosterone is not efficacious in the treatment of subjective symptoms in LOH, or d) a combination of the above.

At the time of designing the trial protocol in 2001, a discussion was ongoing about clinical diagnosis and assessment of the therapeutic effect of testosterone treatment in LOH. In defining our selection criteria for our trial, we largely followed the recommendations for clinical practice based on the 1st Annual Andropause Consensus Meeting, organized by the Endocrine Society in 2000. Because it was not yet clear who might benefit from testosterone treatment, it was decided to keep the main inclusion criteria for this exploratory trial broad and close to clinical practice: subjects with symptoms indicative of LOH in combination with a single assessment of a calculated free testosterone level below the normal range for a young population. Two screening questionnaires for LOH were available (20, 26) as well as the Aging Males’ Symptoms (AMS) rating scale that claimed to measure 1) symptoms of male aging, 2) changes over time, and 3) severity of symptoms (23). In the absence of other relevant symptom questionnaires for use in the LOH population, the AMS rating scale was selected as the primary outcome measure in our study. Although the authors stated that it was unclear whether the AMS scale had relevance for diagnosis of LOH, the AMS rating scale became often used and was culturally and linguistically validated in different languages (27). However, most subsequent studies failed to find an association between total AMS score and testosterone levels (28–32), whereas, weak correlations were reported between AMS scale sub-domains (psychological, somatic, sexual) and testosterone levels (32, 33). Moreover, the validity of the AMS rating scale as an instrument to measure clinical efficacy is unknown since the tool has not been evaluated in a randomized clinical trial in LOH, although it has been empirically demonstrated in open-label studies of truly hypogonadal men (34, 35).

In our study population, the AMS rating scale did not detect alterations in HRQoL related to changes in circulating testosterone levels in LOH when compared with placebo. The characteristics and limitations of the AMS rating scale should be carefully considered before using it as a measure in a clinical intervention trial.

With aging, total testosterone levels decline and SHBG levels increase, leading to even lower non-SHBG bound (free) testosterone levels. We therefore used free testosterone rather than total testosterone as an inclusion criterion in our study, with the lower limit of the normal range of a young, healthy population as a cut-off. In our trial population, the mean baseline SHBG level was relatively high (mean 44.0 nmol/l), with a mean serum total testosterone level at baseline of 14.9 nmol/l. Although the normative ranges for total and free testosterone levels may vary among laboratories and assays, it is generally agreed that the cut-off level for a biochemical diagnosis of hypogonadism is below 8–12 nmol/l for total testosterone (1, 2) and below 0.14–0.31 nmol/l for free testosterone (2). Therefore, we can conclude that our study population might be considered hypogonadal based on calculated free testosterone levels, but had a ‘low-normal’ rather than a hypogonadal status based on total testosterone. The consequence of the low-normal gonadal status of our study population may have been that, for certain symptoms, the baseline testosterone level was above the threshold level for these symptoms and hence increasing serum testosterone levels may not have resulted in
an improvement of symptoms. With respect to the biochemical inclusion criterion for our trial, it is recognized that the calculated free testosterone measure includes the built-in errors of both the testosterone and SHBG assays. In view of this, two screening samples instead of only one would have provided a more robust assessment of testosterone deficiency. In terms of the correct cut-off level, there is still debate ongoing whether to use the normal range of older asymptomatic men instead of that of young, healthy men.

Only the presence of symptoms based on a positive ADAM questionnaire was an inclusion criterion, not symptom severity. In our study population, the average AMS score at baseline ranged from 37.3 to 39.8 points, and was therefore only of ‘mild-to-moderate’ severity (23). As a consequence, there was little room for symptom improvement during treatment and, again, the response to treatment might have been better had only patients with more severe symptoms been included. This is in line with the results from a recent (not yet published) placebo-controlled study with a testosterone gel preparation in symptomatic LOH, where an improvement of the AMS total score was reported after 6 months of treatment (Bouloix et al. Poster presented at the 6th World Congress on the Aging Male, February 21–24, 2008, Tampa, FL (USA)). Compared to our study, the baseline total AMS score was higher (47 vs 38 points in our trial) with larger decreases in the mean total AMS score at month 6 compared with baseline (−10.8 in the testosterone group versus −6.9 in the placebo group). Therefore, the AMS score as measured after 6 months of testosterone gel treatment was in the same range as the baseline AMS score in our study. Furthermore, the number of subjects per group was larger (~180 subjects versus 80 subjects in our trial) and subjects were allowed to adjust their testosterone dose in case of insufficient efficacy.

Another complicating factor is that LOH symptoms such as lack of energy, depressed mood, and sleep problems are generally non-specific and could potentially be caused by other factors such as confounding illnesses which are often present in the older population. When there is no link between the symptoms that men experience and their testosterone status, treatment with a testosterone preparation is not expected to improve these symptoms. In our trial, a large percentage of men could not participate because they did not have a free testosterone level below the cut-off, although they presented with LOH symptoms, indicating the complex relation between symptoms and testosterone levels. The fact that this relationship is not unambiguous is also illustrated by data from other studies investigating patient-reported outcomes in LOH. Although usually favorable trends were seen similar to this study, in some studies, these also did not reach statistical significance (4, 36, 37), whereas in other studies statistically significant improvements of subjective symptoms have been reported (7, 9, 38). It must be emphasized that these studies were generally focused on objective clinical endpoints (hormone levels, bone mineral density, lean body mass, body fat mass, hematocrit) and therefore not powered to demonstrate statistically significant differences in subjective, patient-reported symptoms. When looking at the AMS sub-domains, psychological and somatic symptoms did not improve, whereas sexual symptoms improved only in the oral TU 160 mg/day group. The improvement of sexual function corroborated with an improvement on the DISF-SR which is a validated sexual function rating scale. The results of this and other studies suggest sexual function responds earlier to testosterone treatment in LOH than other aspects of HRQoL, which is in line with the known actions of testosterone. Finally, a combination of the above-mentioned reasons might explain the absence of a significant effect of testosterone treatment as compared with placebo on the AMS score in this study.

Safety

The study showed that treatment with oral TU did not give rise to any safety concern. The preparation was generally well-tolerated with a safety profile comparable with placebo, except for back pain which was reported consistently more frequent in each of the active treatment groups compared with placebo and conjunctivitis which was only reported in the placebo group and not in any of the active treatment groups. A point of caution, as with all androgen preparations, is an increased risk of polycythemia (hematocrit >54%), especially in patients treated with high testosterone doses. In our study, it was demonstrated that a clinically relevant elevation of hematocrit may occur in ~6.5% of subjects using oral TU 240 mg/day. It is therefore recommended that blood hematocrit is monitored, especially during the first year of treatment. If polycythemia develops, the testosterone treatment should be discontinued and resumed at a lower dose when hematocrit has normalized (1, 2). Similarly, our study has also demonstrated that treatment with oral TU does not result in an increased risk of prostate symptoms as evidenced by the absence of clinically relevant effects on PSA or IPSS. Our findings are in line with a recent meta-analysis in which it was reported that there appears to be no relation between serum testosterone levels and risk of prostate cancer (39). Obviously, our study was not powered to investigate the effect of treatment on hard end points such as risk of benign prostate hyperplasia or prostate cancer. It has been theorized that, as a consequence of the use of castor oil as a solvent for TU in the capsules, treatment with oral TU might induce diarrhea. The low prevalence of diarrhea reported in this study of 19 out of 316 subjects (6.0%) suggests that treatment with oral TU may not result in an increased risk of diarrhea or other gastrointestinal symptoms.
Further comparison to other trials and recommendations

In terms of theoretical man-years (MY) of exposure, the present trial (322 subjects for 1 year equals 322 MY) is comparable with the landmark study by Snyder et al. (4) (108 subjects for 3 years equals 324 MY), making it one of the largest studies ever designed in this field; few other studies with a theoretical exposure of > 100 MY are the studies by Amory et al. (210 MY) (10), Nair et al. (174 MY) (13), Emmelot-Vonk et al. (119 MY) (14) and Steidle et al. (102 MY) (7). Our study was well-designed, comparing three different doses of oral TU in a randomized, placebo-controlled setting. With a few exceptions (7, 12), most other studies investigating the effects of testosterone in LOH compared only one dose of testosterone with placebo. The patient population in our trial adequately reflected the typical patient presenting with a request for testosterone treatment. Such patients often present without overt hypogonadism, but rather with non-specific symptoms possibly indicative of LOH, in combination with low-normal serum testosterone levels. For clinical practice, well-controlled efficacy and safety data are essential in this patient population.

The mean baseline total testosterone level of 14.9 nmol/l in our population was lower than the 17.0 nmol/l reported in another study with oral TU (8), but was higher as compared with the baseline levels reported in most other studies in LOH which were in the range of 8.2–13.5 nmol/l (5–7, 10, 13, 14, 40, 41). All studies had different inclusion criteria with respect to baseline testosterone level: in some studies, inclusion was based on a measure of free baseline testosterone level: in some studies, inclusion criteria with respect to relevant LOH signs and symptoms. The variety of inclusion criteria in different studies demonstrates that there is a need for a uniform and valid definition of LOH, not only biochemically but also with respect to relevant LOH signs and symptoms.

In conclusion, oral testosterone supplementation in aging men with borderline testosterone levels and mild-to-moderate symptoms indicative of LOH does not result in a clinically meaningful improvement of psychological and somatic symptoms, whereas a modest improvement may be expected on sexual symptoms. Several recommendations for research can be made. First, a more detailed and uniform definition of LOH is needed, including diagnostic tools. Second, a validated instrument for measurement of testosterone induced changes in HRQoL should be developed in the target population. Third, to make optimal assessment of efficacy and safety of testosterone treatment, well-designed and controlled trials and meta-analyses are needed.

Study 43203 Investigators

The Study 43203 Investigators were: Dr A E Jungwirth, St Johannes Spital, Salzburg (Austria), Prof. J-M Kaufman, University Hospital Gent, Gent (Belgium), Prof. D Vandekerckhove, University Hospital Leuven, Leuven (Belgium), Prof. J J Legros, University Hospital of Liege, Liege (Belgium), Dr C Sigg, Speziali Andrologie, Zürich (Switzerland), Prof. W B Schill, Justus-Liebig University, Giessen (Germany), Dr F M Köhn, Technical University Munich, Munich (Germany), Dr F Giuliano, University Hospital Bicêtre, Le Kremlin-Bicêtre (France), Dr E J H Meuleman, Radboud University Medical Center, Nijmegen (The Netherlands), Dr H P F Koppeschaar, University Medical Center, Utrecht (The Netherlands), Dr L M T de Bie, JUSTUS Medical Expertise, Eindhoven (The Netherlands), Prof. P M G Bouloux, Royal Free Hospital, London (UK), Prof. R Ross, The Northern General Hospital, Sheffield (UK), Dr M Hamilton, University of Aberdeen, Aberdeen (UK).

Declaration of interest

No other potential conflict of interest relevant to this article was reported.

Funding

Financial disclosures: This study (Protocol # 43203) was funded by Schering-Plough. J M H E, M J G H K, and T B P G are employees of Schering-Plough, sponsor of this study. They did not receive any compensation outside their normal salaries. P M G B, J J L, and E J M have received lecture fees from the sponsor. Role of the sponsor: Schering-Plough provided funding, contributed to the study design, provided study medication, and placebo, oversaw quality control at the clinical centers including periodic site visits, and edited the data collected by the centers. The database containing the findings of the individual investigator sites was maintained by the sponsor, and statistical analyses were performed by the sponsor.

Author contribution statement

J M H E was responsible for the conception, design, and execution of the study. J M H E and M J G H K contributed to the statistical analysis and interpretation of the results and critically reviewed the manuscript. T B P G contributed to the interpretation of the results and drafted the manuscript. Study investigators P M G B, J J L, and E J M recruited and treated patients with trial medication, collected study results, contributed to the interpretation of the results, and critically reviewed the manuscript. All authors have approved the manuscript for important intellectual content and are in the position to take public responsibility for this manuscript.

Acknowledgements

The authors thank the investigators and staff of study 43203 for their assistance in conducting this study and the study subjects for their willing participation in this research. The authors would also like to thank Marcel Hekking at Schering-Plough for his contribution to the reporting of this study.

www.eje-online.org
References


Received 5 February 2009
Accepted 9 February 2009