The Effect of Testosterone Replacement Therapy on Nonalcoholic Fatty Liver Disease in Older Hypogonadal Men+

Hae Seung Lee*[1], Sang Hun Han*[1], Ronald Swerdloff[1,2], Youngju Pak[1,2], Matthew Budoff[3], Christina Wang[1,2,4]

Divisions of Endocrinology[1] and Cardiology[3], Department of Medicine and the Clinical and Translational Research Institute[2], The Lundquist Institute and Harbor-UCLA Medical Center, Torrance, CA

(+Oral presentation at Endo 2023)

*These authors contributed equally to the research project)

Keywords: Nonalcoholic fatty liver; male reproductive health; Aging

Corresponding author Christina Wang, MD, Clinical and Translational Science Institute, The Lundquist Institute at Harbor-UCLA Medical Center, 1124 Carson Street, Torrance CA 90502, USA. Email wang@lundquist.org ORCID 0000-0002-8231-792X

Disclosure Summary: The authors have no relevant financial or non-financial interests to disclose

Abstract 251 words

Context. Male hypogonadism is associated with visceral obesity and the metabolic syndrome: factors important for the development of nonalcoholic fatty liver disease (NAFLD). The Testosterone Trials (The T Trials) showed testosterone (T) treatment compared to placebo in older hypogonadal men was associated with decreases in cholesterol and insulin levels suggesting that T treatment may improve NAFLD.

Objective. Compare effects of T versus placebo treatment on NAFLD scores and liver scans in elderly hypogonadal men.

Methods. Secondary data analyses from 479 older hypogonadal men with total T <275 ng/dL from The T Trials were performed. Three clinical liver fat scores: lipid accumulation product (LAP) index, hepatic steatosis index (his), nonalcoholic fatty liver disease-metabolic syndrome (NAFLD-MS) score, and liver computed tomography (CT) Hounsfield unit (HU) and liver to spleen ratio (LSR) were evaluated at baseline and 12 months after treatment.

Results. There were no statistically significant differences of change in LAP index (p=0.98), HSI (p=0.67), and NAFLD-MS (p=0.52) in 246 men treated with T compared to 233 treated with placebo for 12 months. Liver CT showed no statistically significant difference of change in HU (p=0.24; n=71 for T, n=69 for placebo) and LSR (p=0.74; n=55 for T, n=62 for placebo) between the two groups.

Conclusions. Our study did not show improvement of NAFLD in older hypogonadal men after 12 months of T versus placebo treatment, as assessed by three clinical scores and liver CT for hepatic steatosis. Future studies with longer treatment

duration and additional NAFLD diagnostic modalities as primary outcome are warranted.

3

Introduction

Nonalcoholic fatty liver disease (NAFLD) is accumulation of liver fat without excessive alcohol use. NAFLD can progress to nonalcoholic steatohepatitis (NASH) and liver cirrhosis. NAFLD prevalence in adults is estimated to be 25.2% globally (1) and more than 34% in the United States (2,3). NAFLD is strongly associated with obesity, especially visceral adiposity, insulin resistance, and metabolic syndrome. Low testosterone (T) levels in men are associated with increased visceral fat, insulin resistance (4,5) and NAFLD (6,7). Men with low total T had higher hepatic steatosis index than men with higher testosterone but there was no difference in indices of hepatic fibrosis between those with low and higher testosterone concentrations (8). Men with low free T were more likely to have NASH than simple steatosis (9,10). In a rodent study, testosterone treatment ameliorated NAFLD in castrated male rats (11). In a prospective study, T undecanoate injections in 312 hypogonadal men followed for eight years reduced Fatty Liver Index (12). A randomized controlled trial of older men treated for six months with T gel did not alter insulin resistance or liver fat (13). In several studies, men with type 2 diabetes or severe obesity and low T, T treatment improved liver steatosis (14-19). Insufficient data are available on the effects of T treatment on NAFLD in hypogonadal men who do not have metabolic syndrome, type 2 diabetes, or severe obesity. Hence, the objective of this manuscript is to analyze available unpublished data including CT imaging from a large, randomized, placebo-controlled study on hypogonadal older men (The T Trials) to address whether T treatment will improve liver steatosis.

In the testosterone trial (The T Trials), 790 men aged 65 and above were enrolled in T transdermal gel versus placebo treatment for 12 months. Testosterone treatment improved sexual function, mood and depressive symptoms compared to placebo (20). A sub-analyses of The T Trials evaluated the effect of T treatment on cardiovascular biomarkers in the enrolled men (21). Twelve months of T treatment compared to placebo decreased total, high density lipoprotein (HDL)- and low density lipoprotein- (LDL) cholesterol. There were small decreases in fasting insulin and homeostatic model assessment-insulin resistance (HOMA-IR) in the T treatment group, but clinical significance of such small changes in insulin resistance were not clear.

We hypothesized that treatment with T in older hypogonadal men will improve NAFLD after 12 months. Using the data from The T Trials, we performed a secondary analysis of available data and evaluated the prevalence of NAFLD in hypogonadal men and the correlations between the clinical NAFLD scores and liver computed tomography (CT) findings in hypogonadal older men at baseline and after 12-month treatment with transdermal T gel.

Methods

The Testosterone Trials (The T trials)

Our study is a secondary, retrospective analysis of a pre-existing clinical study, called “The T Trials” which were a set of 7 double-blinded, placebo-controlled trials at 12 sites in the United States (20). The data were shared by Peter J Snyder, MD (approved by the University of Pennsylvania), and this secondary analysis study was

approved by the Institutional Review Board of The Lundquist Institute at Harbor-UCLA Medical Center. Hypogonadal men aged ≥ 65 years with serum T concentration ≤ 275 ng/dL received either T gel or placebo for 1 year. Androgel® 1% in a pump bottle was used with initial dose of 5 g gel/day. Serum levels of T at 1, 2, 3, 6, and 9 months were measured and the T gel dose was titrated to maintain a serum T concentration from 400 to 800 ng/dL (the reference range for young men of 19 to 40 years).

Briefly, the inclusion criteria for The T Trials included qualification for at least one of three main trials: sexual functional trial, physical function trial and vitality trial. Exclusion criteria for entry into the trial included: history of prostate cancer; high risk of all prostate cancer (>35%) or high-grade prostate cancer (>7) as per Prostate Risk Calculator; International Prostate Symptom Score greater than 19; and medical conditions known to cause hypogonadism. Additional exclusion factors included: use of medications known to alter serum T concentration; high cardiovascular disease risk (myocardial infarction or stroke within 3 months, unstable angina, New York Heart Association class 3 or 4, systolic blood pressure greater than 180, diastolic greater than 100 mm Hg); hemoglobin A1c > 8.5 % (Hgb A1c); serum creatinine > 2.2 mg/dl; body mass index (BMI) >37 Kg/m^2; and severe depression (PHQ-9 score ≥ 20) (20,21).

Study Participants

A total of 790 subjects were identified who were 65 years and older with serum T level of 275 ng/dL and below. Of these, 705 men completed a follow up period of 12 months. After excluding subjects who did not have baseline and 12-month laboratory data of hepatic panel (AST and ALT), 502 subjects were analyzed. To

ensure absence of effect of alcohol on the liver, 23 men who drank more than 14 drinks per week were additionally excluded, leaving with a total of 479 men. Of these 479 men, 169 men underwent coronary computed tomographic angiography (CCTA) imaging for atherosclerotic plaque trial because this trial began many months after the main T Trials started. Exclusion for CCTA were as follows: estimated glomerular filtration rate <60 ml/min/1.73m^2 or known allergy to iodinated contrast medium; weight >136kg; inability to hold the breath for 10 seconds; a prior diagnosis of tachycardia or irregular heart rhythm; or history of coronary artery bypass graft surgery (22). Among these 169 men, participants were excluded from 12month comparison analysis if subjects did not have both baseline and 12-month CT liver scan data.

Outcome Variables

Medical History:

Baseline history included age, BMI, waist circumference (WC), waist hip ratio (WHR), smoking, alcohol use, ethnicity, diabetes mellitus (DM), hypertension (HTN), hyperlipidemia (HLD), coronary artery disease (CAD), myocardial infarction (MI), stroke, obstructive sleep apnea (OSA), use of cholesterol medication, and use of diabetes medication. BMI and WHR were repeated at 12 months after the treatment.

Laboratory Data:

Baseline and 12-month laboratory tests included AST (aspartate aminotransferase), ALT (alanine aminotransferase), triglyceride (TG), total (TC), HDL-, LDL- cholesterol, Hgb A1c, HOMA-IR, total and free T concentrations. Serum AST and ALT were

measured at Quest Diagnostics as part of the safety laboratory tests (both within and between runs variations was < 3%); TG and Hgb A1c were measured by Cobas Integra 400, Roche Diagnostics (both within and between assay variation was < 3%); and total T and free T were measured by the Brigham Research Assay Core Laboratory and the inter-assay variation was < 7 % for T and <15 % for free T (23).

Nonalcoholic Fatty Liver Disease Score:

Several scoring systems exist for nonalcoholic fatty liver disease scores (24-26). However, based on limited available data, we were able to calculate 3 different scores. First, Lipid Accumulation Product Index (LAP) (27) was calculated using the equation (WC [cm] - 65) * TG [mg/dL]. Cut-off of greater than 30.5 was used for the presence of NAFLD. Second, Hepatic Steatosis Index (HSI) (28) was calculated using the equation 8* ALT/AST ratio + BMI + T2DM (yes = 2, no = 0) + female (yes = 2, no = 0). Cut-off of less than 30 was used for the absence of NAFLD while greater than 36 was used for the presence of NAFLD. Third, Nonalcoholic Fatty Liver Disease in Metabolic Syndrome patients scoring system (NAFLD-MS) (29) was calculated using the equation ALT ≥ 40 U/L (yes = 2, no = 0), AST/ALT ratio ≥ 1 (yes = 1, no = 0), BMI ≥ 25 (yes = 1.5, no = 0), WHR (≥ 0.9 in male and ≥ 0.8 in female; yes = 1, no = 0), T2DM (yes = 1, no = 0). Cut-off of less than 3 was used for absence of NAFLD while 5 or greater was used for presence of NAFLD. All NAFLD scores were calculated for both baseline and 12 months.

Both continuous and categorical variables meeting the cut-off were used in the analysis. The NAFLD scores were calculated using formulae with the Excel® data collection sheet by HSL, then SHH separately calculated the scores using the same

formula. The results from the two investigators were compared and verified by Y. Pak. In addition, if subjects did not report any history of diabetes, for the purpose of calculating these scores, absence of diabetes was assumed.

CT Scan:

Non-enhanced -–CCTA images were obtained for atherosclerotic plaque assessment study of The T Trials participants (22). Two readers blinded to the study groups assessed the scan independently. Hepatic and splenic Hounsfield unit (HU) attenuation was measured in two regions of interest (ROI) in the right liver lobe, and the average value was used as liver HU on non-contrast CT scans. Attenuation of one ROI was used for spleen HU (30). Regions of non-uniform parenchyma attenuation including hepatic vessels were excluded. Liver to spleen ratios (LSR) was calculated by taking the mean HU of both right lobe liver ROIs and dividing by spleen HU measurement. Reproducibility of liver and spleen attenuation measurements were performed on 100 randomly selected scans by the first reader for intra-ready variability. When all 100 scans were read, the second reader performed measurements where both intra- and inter-reader variabilities were calculated. The inter-reader and intra-reader correlation coefficient for liver HU were 0.96 and 0.99 and for spleen HU were 0.99 and 0.99 respectively. The differences between two readers mean HU was 0.4 and 0.1 for liver and spleen respectively. The 95% limits of agreement for inter- and intra- reader measurements for liver HU was -5.63 to 5.25 and -2.79 to 2.52, respectively. The 95% agreement for inter- and intra- reader measurements for spleen HU was -5.68 to 7.15 and -3.9 to 4.42, respectively (30). Two methods were used to indicate the presence of NAFLD. NAFLD liver steatosis was diagnosed with CT criteria if liver HU was less than 40 or

9

if liver to spleen HU ratio (LSR) was less than 1 (31,32). These data were obtained for both baseline and 12 months. Both continuous and categorical variables meeting the cut-off were used.

Statistical Analysis

Data were summarized using mean with standard deviation (SD) for continuous and frequency with percentage for categorical variables. For the comparison of the baseline characteristics between the two groups (T treated vs. placebo), a two-sample t-test and a Chi-square or Fisher's exact test were used for continuous and categorical variables, respectively. The same tests were used to compare changes between the two groups after treatment. For the correlation analyses of LAP index, HSI and NAFLD-MS score with CT findings of NAFLD, Pearson’s correlation coefficient for normally distributed variables or Spearman's correlation coefficient was computed otherwise. To address missing data in the correlation analysis, we employed a list-wise deletion method. This approach involves dropping any man that has a missing value in at least one of the paired or two specified variables. As a result, the sample size (n) for each correlation estimation may vary depending on the specific pairs being analyzed. P-values less than 0.05 were considered statistically significant. All data analyses were carried out using SAS 9.4 (Cary, NC, USA).

Results

Baseline Characteristics:

Baseline characteristics of the subjects in the T group and placebo group did not show statistically significant differences in age, BMI, race, ethnicity, waist-hip ratio (WHR), smoking status, number of drinks per week. Baseline medical history such as type 2 diabetes mellitus, hypertension, coronary artery disease, myocardial infarction, and obstructive sleep apnea (OSA) did not show differences in the two groups (Table 1). More subjects in the T gel group had a history of hyperlipidemia (HLD) than the placebo group (79.92% vs 71.30%, p= 0.03), but the number of subjects taking cholesterol or diabetes medications did not show statistically significant differences between the two groups. Lipid panel, AST, ALT, Hemoglobin A1c, and HOMA-IR, did not show difference at baseline in the two groups (Table 1).

NAFLD scores at baseline:

The three scores to assess NAFLD in the T group and placebo group did not show statistically significant differences (Table 1). The prevalence of NAFLD defined by LAP index greater than 30.5 was 89.02% in the T group and 90.56% in the placebo group (p=0.65). The prevalence of NAFLD defined by HSI equal or greater than 36 showed similar prevalence of 88.62% in the T group and 89.27% in the placebo group (p=0.88). When prevalence was assessed with NAFLD-MS score (NAFLD-MS score equal or greater than 5), the prevalence was 4.07% in the T group and 3.86% in the placebo group (p=1.00).

CT findings at baseline:

The average liver HU was 45.33±12.47 in the testosterone group and 45.37±12.85 in the placebo group and was not significantly different (p=0.98) (Table 1). The liver spleen ratio (LS ratio) was also similar in both groups. The prevalence of NAFLD defined by liver HU was 32.39% in the T group and 27.54% in the placebo group (p=0.58). The prevalence of NAFLD defined by LSR was 21.82% and 20.97% in the T group and the placebo group, respectively (p=1.00).

Serum total and free T concentration and liver fat score and CT findings:

Serum total and free T concentrations were not different between the groups at baseline (Table 1). Serum total and free T concentrations did not differ between men with NAFLD or without NAFLD as defined by the clinical scores or the CT HU or LSR (Table 2). After 12 months, the increase in total T level were significantly higher in the T vs placebo group (285.5±266.4 vs 1.98±73.57 ng/dL, p<0.0001) (Table 3). The change in free T was significantly higher, in the T compared to placebo group(107.9±108.8 vs 0.67±24.63 pg/ml, p<0.0001).

Comparison of NAFLD parameters at baseline and 12 months in T and Placebo treated groups

Despite significantly higher increase in serum total and free T in the T treated compared to the placebo group, there were no significant differences in BMI, WHR, hemoglobin A1c or HOMA-IR between the two groups (Table 3). The change in prevalence of NAFLD defined by LAP, HSI and NAFLD-MS scores were analyzed in each group. The change in prevalence did not show significant differences after 12 months of treatment between the T and the placebo group (Table 3). Prevalence of subjects with absent /borderline NAFLD defined by HSI or NAFLS-MS score also did

not show significant changes in the T group and placebo group after 12 months. The change in liver HU or LSR did not show significant differences between the T treatment and placebo group at 12 months (Table 3). The prevalence of NAFLD defined by liver HU <40 or liver spleen ratio <1 also did not show significant changes between the two groups (Table 3).

Subgroups analyses in men with NAFLD

We then analyzed the effects of 12-month T treatment only on men who met criteria for NAFLD based on LAP (score greater than>30.5), liver HU (<40), or LSR (<1). Within these three subgroups of men with NAFLD, we evaluated 12-month improvement of clinical scores (LAP, HSI, NAFLD-MS) and CT findings (HU and LSR). Across the three subgroups, our study did not show any statistically significance differences in any NAFLD parameters after 12-month in T or placebo treated groups (data not shown).

Correlation between NAFLD scores and CT:

Correlation analysis between clinical scores of NAFLD and CT finding were done using LAP index, HSI, and NAFLD-MS. There was a significant inverse correlation between LAP index and liver HU (r= -0.20628, p=0.0008), although no significant correlation was found between the LAP index and LSR (r= -0.12709, p=0.0628). There was no significant correlation between HSI or NAFLD-MS with CT findings. The combined clinical scores LAP index, HSI, and NAFLD-MS did not have any correlation with liver fat estimated by HU or LSR with CT scans.

Discussion

The prevalence of NAFLD in the United States in the general population is estimated to be at least 34% (2,3) and increases with aging to about 40% in males above 50 years (33). Low testosterone is linked to hepatic steatosis but not to inflammation or fibrosis (8,34). Other studies have also demonstrated that men with lower free T concentrations were more likely to develop NASH versus simple steatosis (9,10). The prevalence of NAFLD in older hypogonadal men in our study was 22 to 30% when assessed by non-contrast liver HU and LSR criteria, which is comparable to other reports (35,36). However, when NAFLD was defined by LAP index (89%), HSI (89%) and NAFLD-MS score (4%) the prevalence was very different, possibly due to overestimation or underestimation of hepatic steatosis status using different scores with different sensitivity and specificity and some scores are designed for special populations with type 2 diabetes or metabolic syndrome (25,29,37,38).

Many different clinical parameters are used for evaluation of NAFLD. LAP index has a sensitivity and specificity of 94% and 85% (37); HSI has a sensitivity of 93% and a specificity of 92% in patients with type 2 diabetes (28,38); and the sensitivity and specificity of NAFLD-MS are 46.0% and 69.8% for patients with metabolic syndrome (29). Our study excluded men with Hgb A1c > 8.5% and BMI > 37 Kg/m^2, and few men had metabolic syndrome explaining the low prevalence using NAFLD-MS score. None of the three scores measure liver fibrosis. Among the scores for fatty liver, only LAP index showed inverse correlation to liver HU whereas HSI and NAFLD-MS score failed to show correlation with liver HU or LSR. Other score systems exist that may be more reliable for predicting liver fibrosis. The NAFLD fibrosis score includes age, BMI, impaired fasting glucose, AST, ALT, platelet count and serum albumin

level. The fibrosis-4 index (Fib-4) utilizes age, AST, ALT, and platelet count. The NAFLD fibrosis score and Fib-4 score has correlation with liver cirrhosis (25,26,39). Because our study was not designed to assess the effect of T on liver steatosis, platelet counts were not in The T Trial database. We were only able to calculate LAP index, HSI and NAFLD-MS scores.

In addition to the scores described above, findings from CT may provide more direct assessment of liver steatosis. The HU of liver in unenhanced CT is typically slightly higher than spleen but HU is lowered as the triglyceride in the liver increases. The sensitivity and specificity of unenhanced CT to assess ≧ 30% macrovesicular steatosis is 73% and 100% respectively at a cutoff of 42 HU (40,41). An absolute liver attenuation of less than 40HU in enhanced CT is commonly used in clinical practice to make the diagnosis. Limitation of diagnosis of hepatic steatosis with CT is that HU of unenhanced liver can be confounded by other disorders of deposition, such as iron, copper, glycogen, and amiodarone (42). Utilizing more sensitive imaging modality in assessment of NAFLD may show hepatic steatosis not identified by CT scan in future studies. For instance, magnetic resonance imaging-proton density fraction has high sensitivity of 95% and specificity of 92% in diagnosis of hepatic steatosis, whereas transient elastography had a sensitivity of 86% and specificity of 88 % (43).

In contrast to our hypothesis, there were no differences in the change of NAFLD scores and CT findings between the T and the placebo group after 12 months of treatment. Because the initial study population included both NAFLD and non-NAFLD participants, we questioned whether including non-NAFLD subjects weakened the possible T effects. Subgroup analysis on NAFLD subjects based on

three clinical diagnostic criteria and liver HU and LSR, showed that 12-month T replacement did not have statistically significant effects on biomarkers of NAFLD. Our results are similar to a 6-month randomized controlled trial of T gel versus placebo treatment of older men with low T levels did not reduce liver fat using magnetic resonance imaging (13). A long-term prospective study when hypogonadal men were administered T undecanoate injections (1000 mg intramuscular injection every 12 week) for 8 years demonstrated significant decrease in Fatty Liver Index (12). The positive effect on decreasing liver steatosis with T undecanoate injections may be related to the decrease in BMI and visceral fat, the higher T concentration attained, and importantly the longer duration treatment of eight years. In male patients with severe obesity and type 2 diabetes, low serum T was associated with increased prevalence of NAFLD (44,45). Testosterone treatment in men with low testosterone levels and type 2 diabetes or severe obesity reduced hepatic steatosis in some but not all studies (14,15,18,19,46).

In our study HOMA-IR decreased (-0.27±7.42) in the testosterone group but not significantly different from the placebo group that showed no decrease (0.09±5.79) over 12 months. This finding is different from the previous cardiovascular biomarker study of The T Trial that showed testosterone treatment cause very small decreases in fasting insulin (-1.7uIU.mL, p=0.02) and HOMA-IR (-0.6, p=0.03) that might not be clinically significant(15). The difference is likely due to the smaller sample size as our study included only 246 men compared to 394 men in the cardiovascular biomarker study who were treated with T.

Our study is a randomized, placebo-controlled trial of T gel treatment in older men without uncontrolled diabetes and gross obesity (20). The NAFLD clinical score data were derived from 497 men enrolled in The T Trials with 169 men who had liver CT fat scores. The only other study that examined T gel treatment for 6 months in older hypogonadal men without diabetes and obesity had 73 men that also demonstrated no T effect on hepatic fat (13). The main limitation to our study was The T Trials were not designed to assess NAFLD. Therefore, lack of available clinical diagnostic data of NAFLD prevented the use of other NAFLD scores such as Fib 4 scores. Further studies incorporating such scores may be helpful in determining true effects of T in hypogonadal men with NAFLD. Our imaging modality was limited to liver CT obtained during cardiac imaging, magnetic resonance imaging or elastography may provide additional information. The gold standard of diagnosis for NAFLD is liver biopsy which was invasive and not part of The T trials, and hence not available for our secondary analysis study of pre-existing data from a large randomized, placebo controlled clinical trial. Our study was only 12 months in duration; perhaps a longer duration of treatment with T for two or three years would have allowed improvements in visceral fat and NAFLD. Therefore, a randomized, placebo-controlled trial protocol is needed to specifically evaluate T effect on NAFLD in hypogonadal men after longer treatment duration.

Acknowledgement

The authors would like to thank Dr. Peter J Snyder, Professor of Medicine at the Perelman School of Medicine at the University of Pennsylvania, Medical Director at Penn Pituitary Center for providing the data from The T Trials.

Data Availability Statement

17

The complete data set is available from Dr. Peter J. Snyder at the University of Pennsylvania.

18

References

1. Younossi ZM, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M. Global epidemiology of nonalcoholic fatty liver disease-Meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatology*. 2016;64(1):73-84.
2. Younossi ZM, Blissett D, Blissett R, Henry L, Stepanova M, Younossi Y, Racila A, Hunt S, Beckerman R. The economic and clinical burden of nonalcoholic fatty liver disease in the United States and Europe. *Hepatology*. 2016;64(5):1577-1586.
3. Kim D, Kim WR, Kim HJ, Therneau TM. Association between noninvasive fibrosis markers and mortality among adults with nonalcoholic fatty liver disease in the United States. *Hepatology*. 2013;57(4):1357-1365.
4. Nielsen TL, Hagen C, Wraae K, Brixen K, Petersen PH, Haug E, Larsen R, Andersen M. Visceral and subcutaneous adipose tissue assessed by magnetic resonance imaging in relation to circulating androgens, sex hormone-binding globulin, and luteinizing hormone in young men. *Journal of Clinical Endocrinology Metabolism*. 2007;92(7):2696-2705.
5. Seidell JC, Björntorp P, Sjöström L, Kvist H, Sannerstedt R. Visceral fat accumulation in men is positively associated with insulin, glucose, and C-peptide levels, but negatively with testosterone levels. *Metabolism*. 1990;39(9):897-901.
6. Mody A, White D, Kanwal F, Garcia JM. Relevance of low testosterone to non-alcoholic fatty liver disease. *Cardiovasc Endocrinol*. 2015;4(3):83-89.
7. Hermoso DAM, Bizerra PFV, Constantin RP, Ishii-Iwamoto EL, Gilglioni EH. Association between metabolic syndrome, hepatic steatosis, and testosterone

deficiency: evidences from studies with men and rodents. *Aging Male*. 2020;23(5):1296-1315.

8. Polyzos SA, Mousiolis A, Mintziori G, Goulis DG. Nonalcoholic fatty liver disease in males with low testosterone concentrations. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*. 2020;14(5):1571-1577.

9. Sarkar M, Yates K, Suzuki A, Lavine J, Gill R, Ziegler T, Terrault N, Dhindsa S. Low Testosterone Is Associated With Nonalcoholic Steatohepatitis and Fibrosis Severity in Men. *Clin Gastroenterol Hepatol*. 2021;19(2):400-402.e402.

10. Van de Velde F, Bekaert M, Hoorens A, Geerts A, T'Sjoen G, Fiers T, Kaufman JM, Van Nieuwenhove Y, Lapauw B. Histologically proven hepatic steatosis associates with lower testosterone levels in men with obesity. *Asian J Androl*. 2020;22(3):252-257.

11. Nikolaenko L, Jia Y, Wang C, Diaz-Arjonilla M, Yee JK, French SW, Liu PY, Laurel S, Chong C, Lee K, Lue Y, Lee WN, Swerdloff RS. Testosterone replacement ameliorates nonalcoholic fatty liver disease in castrated male rats. *Endocrinology*. 2014;155(2):417-428.

12. Al-Qudimat A, Al-Zoubi RM, Yassin AA, Alwani M, Aboumarzouk OM, AlRumaihi K, Talib R, Al Ansari A. Testosterone treatment improves liver function and reduces cardiovascular risk: A long-term prospective study. *Arab journal of urology*. 2021;19(3):376-386.

13. Huang G, Bhasin S, Tang ER, Aakil A, Anderson SW, Jara H, Davda M, Travison TG, Basaria S. Effect of testosterone administration on liver fat in older men with mobility limitation: results from a randomized controlled trial. *J Gerontol A Biol Sci Med Sci*. 2013;68(8):954-959.

14. Apostolov R, Gianatti E, Wong D, Kutaiba N, Gow P, Grossmann M, Sinclair M. Testosterone therapy reduces hepatic steatosis in men with type 2 diabetes and low serum testosterone concentrations. *World J Hepatol*. 2022;14(4):754-765.
15. Hoyos CM, Yee BJ, Phillips CL, Machan EA, Grunstein RR, Liu PY. Body compositional and cardiometabolic effects of testosterone therapy in obese men with severe obstructive sleep apnea: a randomized placebo-controlled trial. *Eur J Endocrinol*. 2015;173(5):X3.
16. Magnussen LV. Testosterone therapy of men with type 2 diabetes mellitus - a randomized, double-blinded, placebo-controlled study. *Dan Med J*. 2017;64(7).
17. Botha J, Velling Magnussen L, Nielsen MH, Nielsen TB, Højlund K, Andersen MS, Handberg A. Microvesicles Correlated with Components of Metabolic Syndrome in Men with Type 2 Diabetes Mellitus and Lowered Testosterone Levels But Were Unaltered by Testosterone Therapy. *J Diabetes Res*. 2017;2017:4257875.
18. Dhindsa S, Ghanim H, Batra M, Kuhadiya ND, Abuaysheh S, Sandhu S, Green K, Makdissi A, Hejna J, Chaudhuri A, Punyanitya M, Dandona P. 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.
19. Maseroli E, Comeglio P, Corno C, Cellai I, Filippi S, Mello T, Galli A, Rapizzi E, Presenti L, Truglia MC, Lotti F, Facchiano E, Beltrame B, Lucchese M, Saad F, Rastrelli G, Maggi M, Vignozzi L. Testosterone treatment is associated with

reduced adipose tissue dysfunction and nonalcoholic fatty liver disease in obese hypogonadal men. *J Endocrinol Invest*. 2021;44(4):819-842.

20. Snyder PJ, Bhasin S, Cunningham GR, Matsumoto AM, Stephens-Shields AJ, Cauley JA, Gill TM, Barrett-Connor E, Swerdloff RS, Wang C, Ensrud KE, Lewis CE, Farrar JT, Cella D, Rosen RC, Pahor M, Crandall JP, Molitch ME, Cifelli D, Dougar D, Fluharty L, Resnick SM, Storer TW, Anton S, Basaria S, Diem SJ, Hou X, Mohler ER, 3rd, Parsons JK, Wenger NK, Zeldow B, Landis JR, Ellenberg SS. Effects of Testosterone Treatment in Older Men. *N Engl J Med*. 2016;374(7):611-624.

21. Mohler ER, 3rd, Ellenberg SS, Lewis CE, Wenger NK, Budoff MJ, Lewis MR, Barrett-Connor E, Swerdloff RS, Stephens-Shields A, Bhasin S, Cauley JA, Crandall JP, Cunningham GR, Ensrud KE, Gill TM, Matsumoto AM, Molitch ME, Pahor M, Preston PE, Hou X, Cifelli D, Snyder PJ. The Effect of Testosterone on Cardiovascular Biomarkers in the Testosterone Trials. *J Clin Endocrinol Metab*. 2018;103(2):681-688.

22. Budoff MJ, Ellenberg SS, Lewis CE, Mohler ER, 3rd, Wenger NK, Bhasin S, Barrett-Connor E, Swerdloff RS, Stephens-Shields A, Cauley JA, Crandall JP, Cunningham GR, Ensrud KE, Gill TM, Matsumoto AM, Molitch ME, Nakanishi R, Nezarat N, Matsumoto S, Hou X, Basaria S, Diem SJ, Wang C, Cifelli D, Snyder PJ. Testosterone Treatment and Coronary Artery Plaque Volume in Older Men With Low Testosterone. *JAMA*. 2017;317(7):708-716.

23. Cunningham GR, Stephens-Shields AJ, Rosen RC, Wang C, Bhasin S, Matsumoto AM, Parsons JK, Gill TM, Molitch ME, Farrar JT, Cella D, Barrett-Connor E, Cauley JA, Cifelli D, Crandall JP, Ensrud KE, Gallagher L, Zeldow B, Lewis CE, Pahor M, Swerdloff RS, Hou X, Anton S, Basaria S, Diem SJ,

Tabatabaie V, Ellenberg SS, Snyder PJ. Testosterone Treatment and Sexual Function in Older Men With Low Testosterone Levels. *J Clin Endocrinol Metab*. 2016;101(8):3096-3104.

24. Vilar-Gomez E, Chalasani N. Non-invasive assessment of non-alcoholic fatty liver disease: Clinical prediction rules and blood-based biomarkers. *J Hepatol*. 2018;68(2):305-315.

25. Saokaew S, Kositamongkol C, Charatcharoenwitthaya P, Srivanichakorn W, Washirasaksiri C, Chaiyakunapruk N, Phisalprapa P. Comparison of noninvasive scoring systems for the prediction of nonalcoholic fatty liver disease in metabolic syndrome patients. *Medicine (Baltimore)*. 2020;99(50):e23619.

26. Lee J, Vali Y, Boursier J, Spijker R, Anstee QM, Bossuyt PM, Zafarmand MH. Prognostic accuracy of FIB-4, NAFLD fibrosis score and APRI for NAFLD-related events: A systematic review. *Liver Int*. 2021;41(2):261-270.

27. Bedogni G, Kahn HS, Bellentani S, Tiribelli C. A simple index of lipid overaccumulation is a good marker of liver steatosis. *BMC Gastroenterol*. 2010;10:98.

28. Lee JH, Kim D, Kim HJ, Lee CH, Yang JI, Kim W, Kim YJ, Yoon JH, Cho SH, Sung MW, Lee HS. Hepatic steatosis index: a simple screening tool reflecting nonalcoholic fatty liver disease. *Dig Liver Dis*. 2010;42(7):503-508.

29. Saokaew S, Kanchanasuwan S, Apisarnthanarak P, Charoensak A, Charatcharoenwitthaya P, Phisalprapa P, Chaiyakunapruk N. Clinical risk scoring for predicting non-alcoholic fatty liver disease in metabolic syndrome patients (NAFLD-MS score). *Liver Int*. 2017;37(10):1535-1543.

30. Zeb I, Li D, Nasir K, Katz R, Larijani VN, Budoff MJ. Computed tomography scans in the evaluation of fatty liver disease in a population based study: the multi-ethnic study of atherosclerosis. *Acad Radiol*. 2012;19(7):811-818.
31. Speliotes EK, Massaro JM, Hoffmann U, Foster MC, Sahani DV, Hirschhorn JN, O'Donnell CJ, Fox CS. Liver fat is reproducibly measured using computed tomography in the Framingham Heart Study. *J Gastroenterol Hepatol*. 2008;23(6):894-899.
32. Schwenzer NF, Springer F, Schraml C, Stefan N, Machann J, Schick F. Non-invasive assessment and quantification of liver steatosis by ultrasound, computed tomography and magnetic resonance. *J Hepatol*. 2009;51(3):433-445.
33. Estes C, Razavi H, Loomba R, Younossi Z, Sanyal AJ. Modeling the epidemic of nonalcoholic fatty liver disease demonstrates an exponential increase in burden of disease. *Hepatology*. 2018;67(1):123-133.
34. Dayton KA, Bril F, Barb D, Lai J, Kalavalapalli S, Cusi K. Severity of non-alcoholic steatohepatitis is not linked to testosterone concentration in patients with type 2 diabetes. *PLoS One*. 2021;16(6):e0251449.
35. Zeb I, Katz R, Nasir K, Ding J, Rezaeian P, Budoff MJ. Relation of nonalcoholic fatty liver disease to the metabolic syndrome: the Multi-Ethnic Study of Atherosclerosis. *J Cardiovasc Comput Tomogr*. 2013;7(5):311-318.
36. Wells MM, Li Z, Addeman B, McKenzie CA, Mujoomdar A, Beaton M, Bird J. Computed Tomography Measurement of Hepatic Steatosis: Prevalence of Hepatic Steatosis in a Canadian Population. *Can J Gastroenterol Hepatol*. 2016;2016:4930987.

37. Ebrahimi M, Seyedi SA, Nabipoorashrafi SA, Rabizadeh S, Sarzaeim M, Yadegar A, Mohammadi F, Bahri RA, Pakravan P, Shafiekhani P, Nakhjavani M, Esteghamati A. Lipid accumulation product (LAP) index for the diagnosis of nonalcoholic fatty liver disease (NAFLD): a systematic review and meta-analysis. *Lipids in Health and Disease*. 2023;22(1):41.
38. Fennoun H, Mansouri SE, Tahiri M, Haraj NE, Aziz SE, Hadad F, Hliwa W, Badr W, Chadli A. Interest of hepatic steatosis index (HSI) in screening for metabolic steatopathy in patients with type 2 diabetes. *Pan Afr Med J*. 2020;37:270.
39. Patel K, Sebastiani G. Limitations of non-invasive tests for assessment of liver fibrosis. *JHEP Rep*. 2020;2(2):100067.
40. Park SH, Kim PN, Kim KW, Lee SW, Yoon SE, Park SW, Ha HK, Lee MG, Hwang S, Lee SG, Yu ES, Cho EY. Macrovesicular hepatic steatosis in living liver donors: use of CT for quantitative and qualitative assessment. *Radiology*. 2006;239(1):105-112.
41. Chartampilas E. Imaging of nonalcoholic fatty liver disease and its clinical utility. *Hormones (Athens)*. 2018;17(1):69-81.
42. Fritz GA, Schoellnast H, Deutschmann HA, Wiltgen M, Brader P, Berghold A, Groell R. Density histogram analysis of unenhanced hepatic computed tomography in patients with diffuse liver diseases. *J Comput Assist Tomogr*. 2006;30(2):201-205.
43. Jia S, Zhao Y, Liu J, Guo X, Chen M, Zhou S, Zhou J. Magnetic Resonance Imaging-Proton Density Fat Fraction vs. Transient Elastography-Controlled Attenuation Parameter in Diagnosing Non-alcoholic Fatty Liver Disease in

Children and Adolescents: A Meta-Analysis of Diagnostic Accuracy. *Frontiers in pediatrics*. 2021;9:784221.

44. Zhang X, Xiao J, Liu Q, Ye Y, Guo W, Cui J, He Q, Feng W, Liu M. Low Serum Total Testosterone Is Associated with Non-Alcoholic Fatty Liver Disease in Men but Not in Women with Type 2 Diabetes Mellitus. *Int J Endocrinol*. 2022;2022:8509204.

45. Yang LJ, Zhou JZ, Zheng YF, Hu X, He ZY, Du LJ, Gu X, Huang XY, Li J, Li YQ, Pan LY, Zhang XX, Gu XJ. Association of non-alcoholic fatty liver disease with total testosterone in non-overweight/obese men with type 2 diabetes mellitus. *J Endocrinol Invest*. 2023.

46. Magnussen LV, Andersen PE, Diaz A, Ostojic J, Højlund K, Hougaard DM, Christensen AN, Nielsen TL, Andersen M. MR spectroscopy of hepatic fat and adiponectin and leptin levels during testosterone therapy in type 2 diabetes: a randomized, double-blinded, placebo-controlled trial. *Eur J Endocrinol*. 2017;177(2):157-168.

Tables

Table 1. Baseline characteristics of testosterone and placebo group.

	Testosterone (n=246)	Placebo (n=233)	P value
Age (years)	72.02±6.01 (n=240)	72.37±5.81 (n=215)	P=0.52
Body Mass Index (kg/m^2)	31.18±3.49	31.09±3.55	P=0.79

	(n=246)	(n=233)	
Waist to hip ratio	1.00±0.05 (n=246)	1.01±0.05 (n=233)	P=0.34
Smoking	17/240 (7.08%)	18/216 (8.33%)	P=0.72
Alcohol (drinks/week)	2.55±3.65 (n=240)	2.98±4.34 (n=215)	P=0.25
White	213/240 (88.75%)	189/216 (87.5%)	P=0.77
Hispanic	12/239 (5.02%)	189/216 (87.5%)	P=0.48
Medical History			
Type 2 Diabetes Mellitus	96/240 (40%)	81/215 (37.67%)	P=0.63
Hypertension	173/240 (72.08%)	152/215 (70.70%)	P=0.75
Hyperlipidemia	191/239 (79.92%)	154/216 (71.30%)	P=0.03
Coronary artery disease	64/240 (26.67%)	55/215 (25.58%)	P=0.83
History of Myocardial Infarction	33/240 (13.75%)	33/215 (15.35%)	P=0.68
History of Stroke	8/240 (3.33%)	10/216 (4.63%)	P=0.63
History of obstructive sleep apnea	52/240 (21.67%)	47/216 (21.76%)	P=1.00
Cholesterol medications	174/240 (72.5%)	139/216 (64.35%)	P=0.06
Diabetes medications	85/240 (35.42%)	78/216 (36.11%)	P=1.00
Labs			
AST (units/L)	24.56±11.06 (n=246)	24.08±11.09 (n=233)	P=0.63
ALT (units/L)	23.73±10.30 (n=246)	23.09±11.84 (n=233)	P=0.52
Triglycerides (mg/dL)	154.4±177.8 (n=242)	156.7±84.1 (n=228)	P=0.85
Total Cholesterol (mg/dL)	162.5±38.02 (n=242)	168.4±34.62 (n=228)	P=0.08
	Testosterone (n=246)	Placebo (n=233)	P value

27

HDL-Cholesterol (mg/dL)	44.52±12.90 (n=242)	44.79±15.14 (n=228)	P=0.83
LDL-Cholesterol (mg/dL)	87.90±1.85 (n=236)	92.67±30.51 (n=222)	P=0.08
Hemoglobin A1c	6.26±0.77 (n=245)	6.30±0.83 (n=229)	P=0.54
HOMA-IR	5.85 ±6.55 (n=246)	5.46±5.59 (n=233)	P=0.48
Total T (ng/dL)	232.6±62.59 (n=245)	237±62.86 (n=231)	P=0.44
Free T (pg/mL)	61.07±19.95 (n=239)	64.72±22.05 (n=228)	P=0.06
NAFLD scores			
LAP (continuous)	78.32±75.18 (n=246)	81.10±47.76 (n=233)	P=0.62
LAP > 30.5 (categorical-positive)	219/246 (89.02%)	211/233 (90.56%)	P=0.65
HSI (continuous)	40.83±4.58 (n=246)	40.71±3.81 (n=233)	P=0.75
HSI ≥ 36 (categorical-positive)	218/246 (88.62%)	208/233 (89.27%)	P=0.88
NAFLD-MS (continuous)	3.52±0.88 (n=246)	3.49±0.80 (n=233)	P=0.67
NAFLD-MS ≥ 5 (categorical-positive)	10/246 (4.07%)	9/233 (3.86%)	P=1.00
CT scores			
Liver HU (continuous)	45.33±12.47 (n=71)	45.37±12.85 (n=69)	P=0.98
Liver HU < 40 (categorical-positive)	23/71 (32.39%)	19/69 (27.54%)	P=0.58
LSR (continuous)	1.29±0.72 (n=55)	1.40±2.07 (n=62)	P=0.70
LSR < 1 (categorical-positive)	12/55 (21.82%)	13/62 (20.97%)	P=1.00

Mean ± SD for continuous and Frequency (%) for categorical variables were reported. Only those without missing data were included for the purpose of statistical analysis. HOMA-

IR=Homeostatic Model Assessment for Insulin Resistance, T=testosterone, LAP=lipid accumulation product index, HSI=hepatic steatosis index, NAFLD-MS=non-alcoholic fatty liver disease in metabolic syndrome patients scoring system, HU=Hounsfield unit, LSR=liver to spleen ratio.

29

Table 2. Serum total and free T concentrations in men categorized by each clinical or CT score with NAFLD (+) or no NAFLD (-)

	Serum Total T (ng/dL)			Serum Free T (ng/dL)		
	NAFLD +	P		NAFLD+	NAFLD-	p
LAP index	233.7±62.94 (n=430)	244.8±60.12 (n=46)	P=0.25	62.78±20.64 (n=422)	63.55±24.94 (n=45)	P=0.82
HSI index	233.2±62.81 (n=423)	246.9±61.00 (n=53)	P=0.14	62.41±20.85 (n=415)	66.39±22.59 (n=52)	P=0.20
NAFLD_MS score	238.8±60.44 (n=19)	234.6±62.85 (n=457)	P=0.76	63.42±15.93 (n=18)	62.83±21.26 (n=449)	P=0.90
Liver HU	242.3±57.97 (n=42)	233.6±58.21 (N=98)	P=0.42	64.93±21.40 (n=41)	61.89±20.78 (n=97)	P=0.44
LSR	246.0±59.62 (n=25)	237.7±57.38 (n=92)	P=0.53	69.35±18.05 (n=24)	61.43±20.45 (n=91)	P=0.09

LAP=lipid accumulation product index, HSI=hepatic steatosis index, NAFLD-MS=non-alcoholic fatty liver disease in metabolic syndrome patients scoring system, HU=Hounsfield unit, LSR=liver to spleen ratio.

Table 3. Change of phenotypic characteristics, laboratory data, NAFLD scores and CT scores from baseline to 12 months in testosterone and placebo group (numerical values).

	Changes(Δ_1) in 12 months in Testosterone (n=246)	Changes(Δ_2)12 months in Placebo (n=233)	Differences of Changes (Δ_1- Δ_2)	P value
Body Mass Index (kg/m^2)	-0.31±1.37 (n=246[1], 246[2])	-0.33±1.27 (n=233[1], 233[2])	+0.0269	P=0.82
Waist to Hip ratio	0.0011±0.05 (n=246[1], 246[2])	-0.0001±0.04 (n=233[1], 233[2])	+0.0012	P=0.77
T (ng/dL)	285.5±266.4 (n=245[1], 242[2])	1.98±73.57 (n=231[1], 231[2])	+283.5	P<0.0001*
Free T (pg/mL)	107.9±108.8 (n=239[1], 235[2])	0.67±24.63 (n=228[1], 225[2])	+107.2	P<0.0001*
Hemoglobin A1c	0.03±0.46 (n=245[1], 221[2])	0.06±0.51 (n=229[1], 211[2])	-0.0658	P=0.16
HOMA-IR	-0.27±7.42 (n=246[1], 246[2])	0.09±5.79 (n=233[1], 233[2])	-0.378	P=0.53
NAFLD scores				
LAP				
LAP (continuous)	-6.1±84.8 (n=246[1], 246[2])	-6.1±39.1 (n=233[1], 233[2])	-0.083	P=0.98
LAP ≥ 30.5 (categorial-positive)	219/246 (89.02%)[1] 203/246 (82.52%)[2]	211/233 (90.56%)[1] 209/233 (89.70%)[2]		P=0.87
HSI				
HSI (continuous)	-0.04±2.17 (n=246[1], 246[2])	0.05±2.25 (n=233[1], 233[2])	-0.09	P=0.67
HSI ≥ 36 (categorical-positive)	218/246 (88.62%)[1] 209/246 (84.96%)[2]	208/233 (89.27%)[1] 209/233 (89.70%)[2]		P=0.20

31

HSI 30-35 (categorical-borderline)	28/246 (11.38%)[1] 37/246 (15.04%)[2]	25/233 (10.73%)[1] 23/233 (9.87%)[2]		P=0.13
	Changes(Δ_1) in 12 months in Testosterone (n=246)	Changes(Δ_2)12 months in Placebo (n=233)	Differences of Changes (Δ_1- Δ_2)	P value
HSI < 30 (categorical-negative)	0/246 (0%)[1] 0/246 (0%)[2]	0/233 (0%)[1] 1/233 (0.43%)[2]		n/a
NAFLD-MS				
NAFLD-MS (continuous)	-0.36±0.96 (n=246[1], 246[2])	-0.31±0.89 (n=233[1], 233[2])	-0.05	P=0.52
NAFLD-MS≥ 5 (categorical-positive)	10/246 (4.07%)[1] 2/246 (0.81%)[2]	9/233 (3.86%)[1] 2/233 (0.86%)[2]		P=0.97
NAFLD-MS 3-4 (categorical-borderline)	173/246 (70.33%)[1] 160/246 (65.04%)[2]	173/233 (74.25%)[1] 149/233 (63.95%)[2]		P=0.24
NAFLD-MS < 3 (categorical-negative)	63/246 (25.61%)[1] 84/246 (34.15%)[2]	51/233 (21.89%)[1] 82/233 (35.19%)[2]		P=0.51
CT scores				
Liver HU (continuous)	0.80±11.15 (n=71[1], 65[2])	-1.49±10.03 (n=69[1], 53[2])	2.29	P=0.24
Liver HU < 40 (categorical-positive)	23/71 (32.39%)[1] 16/65 (24.62%)[2]	19/69 (27.54%)[1] 13/53 (24.53%)[2]		P=0.21
LSR (continuous)	0.17±1.39 (n=55[1], 52[2])	0.27±1.50 (n=62[1], 46[2])	-0.10	P=0.74
LSR < 1 (categorical-positive)	12/55 (21.82%)[1] 11/52 (21.15%)[2]	13/62 (20.97%)[1] 14/46 (30.43%)[2]		P=0.69

Mean ± SD for continuous and Frequency (%) for categorical variables were reported. Only those without missing data were included for the purpose of statistical analysis. [1] Baseline. [2]After 12months treatment.

HOMA-IR=Homeostatic Model Assessment for Insulin Resistance, T=testosterone**,** LAP=lipid accumulation product index, HSI=hepatic steatosis index, NAFLD-MS=non-alcoholic fatty liver disease in metabolic syndrome patients scoring system, HU=Hounsfield unit, LSR=liver to spleen ratio.

33