Title: A low level of serum total testosterone is independently associated with nonalcoholic fatty liver disease

Authors:

Sunmi Kim (arpsm@naver.com)
Hyuktae Kwon (hyuktae@gmail.com)
Jin Ho Park (pihn@snuh.org)
Belong Cho (belong@snu.ac.kr)
Donghee Kim (messmd@chol.com)
Seung Won Oh (drunkenmuse@gmail.com)
Cheol Min Lee (bigbangx@snuh.org)
Ho chun Choi (skyho331@gmail.com)

Version: 4 Date: 29 March 2012

Author's response to reviews: see over
March 29, 2012

Journal : BMC Gastroenterology

Editor : Dr Ali Canbay

Dear Editor

We really appreciated your e-mail of March 13, 2012. We are pleased to hear that our manuscript "A low level of serum total testosterone is independently associated with nonalcoholic fatty liver disease" has a chance to be re-evaluated.

We would also like to thank the reviewers for the thorough examination of our manuscript and excellent comments which have helped us to improve our manuscript. We have addressed the reviewer’s comments below.

We believe that the manuscript has improved substantially thanks to the comments from editor and reviewer, and we hope that it will be acceptable for the journal. We look forward to having our manuscript published in "BMC gastroenterology"

Sincerely yours,

Hyuktae Kwon, MD, MPH

Department of Family Medicine, Healthcare Research Institute, Seoul National University Hospital Healthcare System Gangnam Center, 39th Floor, Gangnam Finance Center 737, Yeoksam-dong, Gangnam-gu, Seoul 135-984, South Korea

Tel.: 82-2-2112-5500 (Fax: 82-2-2112-5510)

E-mail: hyuktae@gmail.com
Responses to the Reviewers’ comments

1. Reviewer’s report: Kim and coworkers present a retrospective cohort study investigating the association of testosterone and non-alcoholic fatty liver disease. This work nicely complements previous studies in diabetes and metabolic syndrome. The present study is the first report on testosterone levels in the setting of fatty liver disease. All data are consistent and adequate presented. Manuscript adheres to the relevant formal standards for reporting and data deposition. Title and abstract accurately convey the contents of the study. However, several shortcomings in the methodology need to be addressed in the manuscript.

comment #1: Most importantly, it has to be mentioned that the gold standard for diagnosis of a non-alcoholic fatty liver disease still is liver biopsy. This approach also allows to differentiate NASH from NAFLD, which has not been considered in the present study. For further analysis of NASH, alternative serological strategies may help to overcome the missing biopsy in the majority of patients.

Response: We agree with the reviewer that the gold standard for NAFLD diagnosis is still liver biopsy. We described that as the first limitation in the discussion section. However, the reason liver biopsy was not taken from the study subjects was that it was not clinically necessary for the healthy participants who voluntarily visited the Hospital for routine health checkups. Ultrasonography instead of biopsy has been largely used for diagnosis of NAFLD [1]. It has also proved its clinical efficacy through many clinical studies regarding NAFLD [2, 3]. Moreover, from an ethical point of view, it is difficult to perform a biopsy for detection of NASH from NAFLD. NAFLD encompasses a wide spectrum of disease ranging from accumulation of fat (fatty liver) to various degrees of inflammation and fibrosis (NASH), and finally to cryptogenic cirrhosis and its clinical sequelae (HCC, liver decompensation) [4]. In this study, we scoped on the NAFLD rather than focusing on NASH. This study was retrospective analysis of clinical data, which was not gathered for the study per se. Therefore, regretfully, we could not gather the alternative serological strategies for NASH such as CK 18(as reviewer commented) or FGF21. Instead, preparing responses to the reviewer's comments, we
applied the NAFLD fibrosis score formula to further analysis of NASH from NAFLD. The NAFLD fibrosis score formula is $\text{NAFLD score} = -1.675 + 0.037 \times \text{age (years)} + 0.094 \times \text{BMI (kg/m}^2\text{)} + 1.13 \times \text{IFG/diabetes (yes = 1, no = 0)} + 0.99 \times \text{AST/ALT ratio} - 0.013 \times \text{platelets (10}^9\text{/L)} - 0.66 \times \text{albumin (g/dL)} [5, 6]. According to the results by Angulo et al. and Wong et al., a score value greater than 0.676 (high cutoff) would predict severe fibrosis, and a score value lower than -1.455 (low cutoff) would exclude severe fibrosis, so liver biopsy could be avoided with high negative predictive value [5, 6]. In this study, only six subjects had the NAFLD fibrosis score > 0.676, which predicts severe fibrosis. The distribution of the NAFLD fibrosis score according to serum total testosterone was as follows:

<table>
<thead>
<tr>
<th>Total testosterone (ng/mL)</th>
<th>NAFLD fibrosis score</th>
<th>Low risk of fibrosis ( &lt; -1.455)</th>
<th>Indeterminate</th>
<th>Predict severe fibrosis ( &gt; 0.676)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st quintile (0.11–3.17)</td>
<td>37 (56.06%)</td>
<td>27 (40.91%)</td>
<td>2 (3.03%)</td>
<td></td>
<td>66</td>
</tr>
<tr>
<td>2nd quintile (3.18–3.88)</td>
<td>26 (46.43%)</td>
<td>29 (51.79%)</td>
<td>1 (1.79%)</td>
<td></td>
<td>56</td>
</tr>
<tr>
<td>3rd quintile (3.92–4.78)</td>
<td>30 (53.57%)</td>
<td>23 (41.07%)</td>
<td>3 (5.36%)</td>
<td></td>
<td>56</td>
</tr>
<tr>
<td>4th quintile (4.80–5.70)</td>
<td>30 (63.83%)</td>
<td>17 (36.17%)</td>
<td>0</td>
<td></td>
<td>47</td>
</tr>
<tr>
<td>5th quintile (5.71–13.43)</td>
<td>15 (57.69%)</td>
<td>11 (42.31%)</td>
<td>0</td>
<td></td>
<td>26</td>
</tr>
<tr>
<td>Total</td>
<td>138 (54.98%)</td>
<td>107 (42.63%)</td>
<td>6 (2.39%)</td>
<td></td>
<td>251</td>
</tr>
</tbody>
</table>

Additionally, we analyzed the association between NAFLD and serum testosterone levels using multiple logistic regression after six subjects with the NAFLD fibrosis score > 0.676 were excluded.
from the original 495 subjects. The results were as follows:

The association between NAFLD and serum testosterone levels in the 489 subjects

<table>
<thead>
<tr>
<th>Total testosterone (ng/mL)</th>
<th>Model 1*</th>
<th></th>
<th>Model 2†</th>
<th></th>
<th>Model 3‡</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1st quintile (0.11–3.17),</td>
<td>4.86 (2.29–10.29)</td>
<td>4.72 (2.21–10.08)</td>
<td>4.27 (1.95–9.36)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2nd quintile (3.18–3.92)</td>
<td>2.51 (1.22–5.17)</td>
<td>2.37 (1.14–4.93)</td>
<td>2.22 (1.04–4.72)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rd quintile (3.93–4.80)</td>
<td>3.21 (1.57–6.55)</td>
<td>3.17 (1.54–6.50)</td>
<td>2.59 (1.23–5.46)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4th quintile (4.81–5.72)</td>
<td>2.30 (1.12–4.73)</td>
<td>2.24 (1.08–4.63)</td>
<td>2.02 (0.95–4.28)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5th quintile (5.75–13.43)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p for the trend < 0.001 < 0.001 0.001

*a* Multiple logistic regression was used to analyze the association between NAFLD and serum total testosterone levels after 6 subjects with the NAFLD fibrosis score > 0.676 were excluded from 495 subjects. A score value greater than 0.676 (high cutoff) would predict severe fibrosis.

*b* The NAFLD fibrosis score was calculated according to the following formula:

\[-1.675 + 0.037 \times \text{age (yr)} + 0.094 \times \text{BMI (kg/m}^2\text{)} + 1.13 \times \text{IFG/diabetes (yes = 1, no = 0)} + 0.99 \times \text{AST/ALT ratio} - 0.013 \times \text{platelet (x10}^9\text{/L)} - 0.66 \times \text{albumin (g/dL)}.

Model 1: adjusted for age, smoking, diabetes, exercise, BMI, TGs, and HDL-C

Model 2: Model 1 + HOMA-IR and hs-CRP

Model 3: Model 2 + VAT (as the continuous variable)

OR, odds ratio; CI, confidence interval

We also analyzed the association between NAFLD and serum testosterone levels using multiple logistic regression after 113 subjects with the NAFLD fibrosis score ≥-1.455 which could not exclude severe fibrosis were subtracted from the original 495 subjects. The results were as follows:

The association between NAFLD and serum testosterone levels in the 382 subjects

<table>
<thead>
<tr>
<th>Total testosterone (ng/mL)</th>
<th>Model 1*</th>
<th></th>
<th>Model 2†</th>
<th></th>
<th>Model 3‡</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1st quintile (0.13–3.26),</td>
<td>5.71 (2.27–14.38)</td>
<td>5.65 (2.22–14.41)</td>
<td>5.69 (2.14–15.12)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2nd quintile (3.27–4.07)</td>
<td>2.28 (0.91–5.72)</td>
<td>2.10 (0.82–5.37)</td>
<td>1.99 (0.75–5.31)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rd quintile (4.10–4.93)</td>
<td>3.53 (1.42–8.77)</td>
<td>3.55 (1.41–8.93)</td>
<td>2.89 (1.10–7.55)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4th quintile (4.95–5.84)</td>
<td>2.32 (0.91–5.92)</td>
<td>2.26 (0.88–5.82)</td>
<td>2.26 (0.84–6.05)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5th quintile (5.87–13.43)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p for the trend 0.001 0.001 0.002*
Multiple logistic regression was used to analyze the association between NAFLD and serum total testosterone levels after 113 subjects with the NAFLD fibrosis score \( \geq -1.455 \) were excluded from 495 subjects. A score value lower than -1.455 (low cutoff) would exclude severe fibrosis.

The NAFLD fibrosis score was calculated according to the following formula:

\[
-1.675 + 0.037 \times \text{age (yr)} + 0.094 \times \text{BMI (kg/m}^2\text{)} + 1.13 \times \text{IFG/diabetes (yes = 1, no = 0)} + 0.99 \times \text{AST/ALT ratio} - 0.013 \times \text{platelet (} \times 10^9/L\text{)} - 0.66 \times \text{albumin (g/dL)}.
\]

Model 1: adjusted for age, smoking, diabetes, exercise, BMI, TGs, and HDL-C

Model 2: Model 1 + HOMA-IR and hs-CRP

Model 3: Model 2 + VAT (as the continuous variable)

OR, odds ratio; CI, confidence interval

The results showed the association between low level of serum total testosterone and NAFLD was significant in the both subgroup excluding the subjects with predictive value of severe fibrosis and subgroup including the subjects with negative predictive value of severe fibrosis by using the NAFLD fibrosis score formula.

comment #2: In sonographic examination rounding of dorsal liver margin should have also been used as a parameter in fatty liver disease. If this has been done, the authors should mention this in the methods section. Also a fibroscan would have been of interest, especially while lacking histology to judge at least advanced fibrosis or cirrhosis versus early disease. Comments of direct or indirect signs of portal hypertension as signs of hepatic cirrhosis resulting in altered testosterone levels are missing.

Response:

On my regrets, we used neither rounding of dorsal liver margin as a parameter in fatty liver disease nor fibroscan. We excluded the subjects with HBsAg or anti-HCV positivity and those drinking heavily, who occupy the majority of the patients with hepatic cirrhosis in Korea [7]. Among the participants finally included, there were no subjects who had the evidence of hepatic cirrhosis on ultrasonography or abdominal CT or of portal hypertension including splenomegaly and esophageal varix found by gastroscopy. (Gastroscopy was performed for all subjects in the routine health checkups of the Seoul National University Hospital Healthcare System Gangnam Center.)
We added the following comments in the **Subjects and methods** (Participants and study design) section:

*Among the participants included, there were no subjects who had the evidence of hepatic cirrhosis or portal hypertension (such as splenomegaly and esophageal varix) on ultrasonography, abdominal CT or gastroscopy.*

**comment #3:** The differential diagnosis of increased echogenicity of liver tissue remains only incompletely addressed. Incomplete clinical data regarding non-infectious reasons of hepatic disease or other reasons for sonographic findings as for example chemotherapy, metabolic or infectious disease are reported. The completeness of this information cannot be found in the tables.

**Response:**

As we described in the Subjects and methods section, we excluded the subject who had the risk of fatty liver due to hepatitis B or C virus by taking serology tests including HBsAg and anti-HCV antibody. In addition, we also excluded the subjects who had drunk more than 140 g of alcohol per week, had a medical history of other types of hepatitis such as autoimmune hepatitis and chronic viral liver disease, cholestasis, and other metabolic liver diseases, or had previously used steatogenic medications including antiretroviral drugs, antiarrhythmic drugs, anticancer drugs, corticosteroids and hormone.

In Korea, the most common cause of liver disease is hepatitis B virus (41-58%), followed by alcohol (26-37%) and hepatitis C virus (7-11%) [7]. The other causes, other than viral hepatitis and alcohol, comprise only a minor proportion, accounting for less than 5% [7]. A nationwide survey for prevalence of rare hepatic diseases (Wilson disease, autoimmune hepatitis) was performed by the Korean Association for the Study of the Liver in 2004. The estimated prevalence of Wilson disease and autoimmune hepatitis were about 0.6/100,000 [8] and less than 0.5% [9], respectively. In case of autoimmune hepatitis, women occupied 90% and men only 10%. Therefore, even though we could not completely exclude the possibility of the rare causes of liver disease, it might not have a
significant effect on the results in this study.

To describe in more detail, we revised the sentences in the Subjects and methods (Participants and study design) section as follows:

By using self-report questionnaires, we also excluded 433 men who consumed more than 140 g of alcohol per week, had a medical history of other types of hepatitis such as autoimmune hepatitis and chronic viral liver disease, cholestasis, and other metabolic liver diseases, or had previously used steatogenic medications including antiretroviral drugs, antiarrhythmic drugs, anticancer drugs, corticosteroids and hormone.

In addition, we added the following sentences to the limitation:

Second, although we checked the medical history of liver disease and previous use of steatogenic drugs through the self-report questionnaires, it was possible that rare liver diseases such as autoimmune hepatitis and Wilson's disease and fatty liver disease caused by steatogenic drugs were not completely excluded considering the recall bias of self-report questionnaires.

**comment #4:** While the authors clearly explain the rationale to correlate testosterone levels with non-alcoholic fatty liver disease, it remains unclear why total testosterone was analysed and why free testosterone and sex-hormone binding globulin was not analysed. From the literature, SHBG would be expected to be elevated in advanced liver disease and cirrhosis thereby resulting in lower levels of free testosterone. To address this important issue, measurement of SHBG and free testosterone should be completed if serum samples of patients are still available.

**Response:**

As mentioned earlier, this study is a retrospective analysis of clinical data of the healthy men who voluntarily took the routine health checkups, which was not gathered for the study per se. Therefore, regrettfully, we could not gather the data of free testosterone and sex-hormone binding globulin and the serum samples of participants are not available at this moment. As we described above, among
the participants included, there were no subjects with the evidence of advanced liver disease or hepatic cirrhosis on the abdominal ultrasonography and CT, serum albumin and serum bilirubin. In addition, only six subjects had the score > 0.676 in the result of estimating the NAFLD fibrosis score.

**Comment #5:** Testosterone levels peak in the early morning and standard time intervals for routine diagnostics usually end at 10 A.M. It remains unclear why the authors decided to exceed this time in favour of increased patient numbers because this approach definitely limits the comparability of testosterone levels.

**Response:**

In this study, we used clinical data from routine health checkups in our hospital, and the room for blood samples was open from 8 a.m. to 11 a.m. However, most of the participants took the blood test in early hours (about 8 a.m. to 9 a.m.).

As reviewer commented, it is widely recommended that testosterone should be assessed in the early morning, before 10 a.m., based on studies showing a significant diurnal variation in testosterone, in which the highest levels were observed in the early morning and nadir levels in the evening hours [10]. However, these studies involved small samples of young men. Older men were reported to have substantial blunting of the diurnal variation in testosterone level [11]. In a large population study in 3,006 middle-aged and older (40 – 94 years) ambulatory men, mean testosterone levels were unchanged from early morning (6 a.m.) until the early afternoon (2 p.m.), and declined only modestly thereafter by 13% [12]. In this study, the mean age was 54.98 years in subjects with NAFLD and 53.85 without NAFLD. Furthermore, 90% (445 of 495) of the subjects were older than 40 years. Therefore, we believe that blood sampling time might have little effects on the results.

**Comment #6:** Laboratory data are supposed to be acquired according to good clinical practice. Why alkaline phophatase and bilirubin are not analysed is not explained. Furthermore, levels of inflammatory cytokines as TNFa and IL-1 and sIL-6r would also have been of interest as remarked in discussion. In the context of inflammation, serum CK 18 as a marker of steatohepatitis in NAFLD would have been of interest. If serum samples of patients are still
available, measurement of CK 18 should be completed.

Response:

According to the reviewer’s suggestion, we added the analyses of alkaline phosphatase and total bilirubin to the table 1 as follows.

<table>
<thead>
<tr>
<th></th>
<th>no NAFLD (n = 244)</th>
<th>NAFLD (n = 251)</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total bilirubin (mg/dL)</td>
<td>1.1 (0.85–1.3)</td>
<td>1.1 (0.9–1.3)</td>
<td>0.687</td>
</tr>
<tr>
<td>ALP (IU/L)</td>
<td>59 (49.5–69)</td>
<td>62 (52–72)</td>
<td>0.018</td>
</tr>
</tbody>
</table>

*Data are presented as the median (range, Quartile 25–Q75) and were tested by the two-sample Wilcoxon rank-sum test

We agree with the reviewer in his/her point that levels of inflammatory cytokines and serum CK 18 as a marker of steatohepatitis in NAFLD would have been of interest. Regretfully, we did not measure them, because there were no tests for them in routine health checkups. In addition, the serum samples of participants were discarded and are not available now. We would consider these markers in the next studies.

comment #7: Why was a CT scan performed in all patients (as a routine examination)? Or was this done for the sole purpose of this study specifically for measurement of VAT? In this case, an ethics approval may be helpful.

Response:

The participants chose the abdominal CT scan for more detailed evaluation of abdominal solid organs of their own accord. In Korea, as the cost of a CT scan is not high, many examinee have taken the CT scan for routine health screening even though they didn't have any symptoms.

We corrected the sentence in the Subjects and methods section as follows in order to clarify the participants took the CT scans of their own free will:

We included only the remaining 1057 men who underwent an examination of serum total testosterone, abdominal ultrasonography, and abdominal CT scans of their own accord, which was necessarily a convenience sample.
2. Reviewer's report: Kim et al. investigated the association of serum total testosterone with non alcoholic fatty liver disease in 495 male probands who underwent abdominal ultrasound, abdominal CT scans (to measure visceral adipose tissue, VAT) and measurement of total serum testosterone. Patients with alcohol consumption >140g/d or hepatitis were excluded from the study. The authors find that probands with NAFLD had significantly lower total testosterone levels. The result is that the association between total testosterone levels and the odds ratio for presence of NAFLD was significant, even after adjusting for potential confounders such as age, BMI, diabetes, HOMA-IR and VAT.

Comments:
The manuscript is well written and the results are clearly presented in three tables and one figure. The results seem valid, are in accordance with previous reports from other groups and are of interest, as it goes beyond previous studies by inclusion of VAT measurement.

Major Revision:
comment #1: Study design: please state whether this cross sectional study was a prospective or retrospective study.

Response:
According to the reviewer’s suggestion, we corrected the sentence in the Subjects and methods section and Design and methods section of Abstract as follows:

This study is a retrospective observational cross-sectional one of healthy Korean men who were aged 20 years or more and who voluntarily visited the Seoul National University Hospital Healthcare System Gangnam Center between January 2008 and April 2010 for routine health checkups.

This study is a retrospective observational cross-sectional one of healthy Korean men and was conducted at the Seoul National University Hospital Healthcare System Gangnam Center.

It is understood that the sample is a convenience sample because only the participants who
underwent measurement of testosterone, ultrasound and abdominal CT scans could be included. The original population was comprised of healthy men visiting the clinic for routine health checkups. It appears that roughly 7200 of 28600 men then received an abdominal CT scan and ultimately 495 of those men could be included in the study. I wonder why a CT scan was ordered in healthy men at a routine checkup? What conditions required a CT and what was found? Did these men then receive a diagnosis from those CTs or the condition leading to the CT and can therefore not be labeled “healthy” anymore?

My concern is: if the CT scan was not a pre-planned diagnostic tool in a prospective study, then including only men that required a CT scan for medical reasons harbours the risk of introducing a strong selection bias. The present study seems to include only those men that were selected because they received a CT scan for medical reasons that are left unexplained. As it is exactly this CT scan that allows measurement of VAT, and that sets this study apart from others, it is absolutely necessary to explain why a CT scan was ordered in the first place. The “Methods” part starts with “This study is ... one of healthy ... men ...”. Is that still true after examining the reasons for and results of the CT scans?

In short, the reader needs to know:

comment #2: for what reason were CT scans ordered?

Response:

CT scans were not forced by the physician but primarily chosen by the participants who wanted more detailed evaluation of abdominal solid organs of their own accord. This is retrospective study, that is to say, CT scans were not imposed for this study but had been performed before designing this study. The cost of CT scans in Korea is far lower than that in USA, therefore, quite a number of examinee have often taken abdominal CT scans for routine health screening to do more detailed evaluations of abdominal solid organs even though they don’t have any symptoms. The examinee who participated in the health checkup program at the Seoul National University Hospital Healthcare System Gangnam Center were informed of the radiation hazard, the possible risks of CT contrast before deciding to take a CT scans and provided informed conscent.
We corrected the sentence in the **Subjects and methods** section as follows in order to clarify the participants took the CT scans of their own free will:

We included only the remaining 1057 men who underwent an examination of serum total testosterone, abdominal ultrasonography, and abdominal CT scans **of their own accord**, which was necessarily a convenience sample.

**comment #3:** what were the main findings in the CT scans of these men, the study population: are they still healthy?

**Response:**
We checked all the findings on the CT scans of the 495 subjects finally included with available data. The subjects had only clinically insignificant findings which included hepatic cysts, hemangioma, hepatic calcification, renal cyst, silent renal stone, silent gallbladder stone and tiny gallbladder polyp. Because such radiologic findings are not believed to be associated with general health status, we regarded them as healthy.

**comment #4:**
for the discussion: do the CT scan results suggest a selection bias?

**Response:**
The selection bias could be suggested in that the subjects more concerned about their health status may be likely to be received the abdominal CT scan more. However, we believe the findings of the CT scans themselves did not result in the selection bias. As mentioned earlier, the findings of the CT scans were clinically insignificant. Furthermore, the study participants could be generally regarded to be representative of the general population, which were presented in the other article using the study population enrolled in a health checkup program at the Seoul National University Hospital Healthcare System Gangnam Center [13].
References


