**Supplementary Materials**

**Supplementary Methods**

***Blood sample processing protocol***

Laboratory analyses were performed in accordance with strict uniform specifications in large central hospital laboratories in four provinces. To assure the accuracy and consistency of testosterone between the partner laboratories, we have formulated a unified process of the corresponding international published specifications.[1]

Fasting blood samples were collected from each participant between 7:00 am and 11:00 am. Blood samples were collected by trained and certified nurses and stored in barcoded vacuum pipettes. The tubes were then scanned with a barcode reader to associate with the participant’s unique identification number and to assure accurately time for the centrifugation of the blood sample. After centrifugation, plasma and buffy coats were immediately separated from the blood samples.

Testosterone and other laboratory indices were measured at the local standard laboratory of the participants. All procedures were performed in accordance with the Standard Operating Procedure, which was formulated by the project team. The measurement of testosterone was performed as follows. (1) A volume of 200 μL of serum with internal standard was incubated for 30 min. (2) Dissociation buffer was added and incubated for 45 min. (3) Dissociate testosterone from binding proteins by adding liquid extraction solution and extraction. (4) The lipid fraction was isolated from the sample by adding deprotonation buffer and extraction. (5) The phospholipids and other polar lipid fractions were removed. (6) The sample was reconstituted. (7) Liquid-chromatography tandem-mass-spectrometry analysis and data collection were performed.

To ensure consistent and credible results from different laboratories, we established corresponding quality control (QC) procedures. Three Bench QC serum samples with different concentrations of total testosterone (TT) from low to high concentrations were used. Each batch of samples was analyzed together with QC samples and blank samples. The following QC criteria were established according to the international common QC principle. If a QC sample result exceeds the mean ± 4 SD, the batch is considered unqualified. If a QC sample result exceeds the mean ±3SD, the batch is considered warning and the processing needs to be reviewed. If the results of two consecutive QC samples exceed the mean value of +/－ 2SD , the batch is considered unqualified. If the difference between the results of two QC samples within a batch exceeds 4SD, the batch is considered unqualified. If 10 consecutive QC sample results are on the same side of the mean, the batch involved is considered unqualified.

***Sensitivity Analyses***

1. We imputed missing values in the study population according to a similar study’s methods.[2] Missing values were input by the median (if continuous) or mode (if categorical) of existing cases of that variable, and the regression analysis was performed again.

2. We further examined the relationship between sleep quality and TT concentrations. The sleep quality was assessed by the Pittsburgh Sleep Quality Index (PSQI), which consists 19 self-rated questions. These questions assess various factors and are grouped into the following seven component scores: subjective sleep quality, sleep latency, sleep duration, habitual sleep efficiency, sleep disturbances, the use of sleeping medications, and daytime dysfunction. The seven component scores are then summed to yield a global PSQI score ranging from 0 to 21.[3] We obtained the PSQI score from the WeCHAT study to identify the participants’ sleep quality. Adjusted multivariable logistic regression analyses were performed for examining the relationship between sleep quality and TT concentrations.

3. A previous study showed that <300 ng/dL is generally used as the TT standard for late-onset hypogonadism (LOH).[4] Therefore, we also performed a sensitivity analysis using a cut-off of 300 ng/dL to examine the association between the sleep duration and the prevalence of LOH.

4. There is currently no uniform cut-off value for a long and short sleep duration. Therefore, we selected some shorter cut-off values for the sensitivity analysis to further ensure that the cut-off value of a long and short sleep duration is reasonable and practical.

5. We performed a propensity score matching of 1:3 to balance the numbers of the two groups to make the long sleep and short sleep groups more comparable and representative. Propensity scores were calculated on the basis of age, BMI, marital status, and the number of chronic diseases.

***Analyze for power of test***

We also performed a post hoc statistical power (1 − β) calculation based on R statistical software (http://www.R-project.org, The R Foundation, Vienna, Austria).



**Supplementary Figure 1:** Detailed inclusion and exclusion process of participants. WeCHAT: West China Health and Aging Trend.



**Supplementary Figure 2:** Non-linear trajectories of TT levels with the changes of sleep duration among participants. There was a significant threshold effect, with insignificant effects to the left of the inflection point, and a significant association of longer sleep duration with higher TT levels to the right of the inflection point. TT: Total testosterone. Red dot: Sleep duration. Blue dot: The upper and lower limits of 95% confidence interval of sleep duration.

**Supplementary Table 1: Baseline characteristics of men over the age of 60 with different sleep durations.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Items  |  | Sleep duration (h) | Statistics | *P*-value |
| **Total****(*n*=1628)** | **<9.5 h****(*n*=1477)** | **≥9.5 h****(*n*=151)** |
| Age (years) | 68.5 ± 5.9 | 68.18 ± 5.85 | 68.95 ± 6.55 | －1.52† | 0.128 |
| BMI (kg/m2) | 24.59 ± 3.72 | 24.95 ± 4.92 | 25.89 ± 7.40 | －2.12† | 0.710 |
| Testosterone (ng/dL) | 428.00 (314.00－554.25) | 427.00 (314.00－549.00) | 442.00 (321.85－580.00) | －1.11‡ | 0.270 |
| Education |  |  |  | 8.32§ | 0.081 |
| Illiteracy | 309 (18.98) | 272 (18.42) | 37 (24.50) |  |  |
| Primary school | 663 (40.72) | 612 (41.44) | 51 (33.77) |  |  |
| Junior high school | 354 (21.74) | 320 (21.67) | 34 (22.52) |  |  |
| High school | 228 (14.01) | 202 (13.68) | 26 (17.22) |  |  |
| Undergraduate and above | 74 (4.55) | 71 (4.81) | 3 (1.99) |  |  |
| Marital |  |  |  | － | 0.941|| |
| Unmarried | 11 (0.68) | 10 (0.68) | 1 (0.66) |  |  |
| Married | 1472 (90.42) | 1337 (90.52) | 135 (89.40) |  |  |
| Divorce | 15 (0.92) | 13 (0.88) | 2 (1.32) |  |  |
| Widowed | 130 (7.98) | 117 (7.92) | 13 (8.61) |  |  |
| Nutrition status |  |  |  | 17.13§ | 0.001 |
| Normal | 1236 (75.92) | 1141 (77.52) | 95 (62.91) |  |  |
| Malnutrition | 352 (21.62) | 304 (20.58) | 48 (31.79) |  |  |
| Unknown | 40 (2.46) | 32 (2.17) | 8 (5.30) |  |  |
| Cognitive performances |  |  |  | － | 0.008|| |
| Normal | 1390 (85.38) | 1274 (86.26) | 116 (76.82) |  |  |
| Impair | 224 (13.76) | 191 (12.93) | 33 (21.85) |  |  |
| Unknown | 14 (0.86) | 12 (0.81) | 2 (1.32) |  |  |
| Physical frailty  |  |  |  | 2.68§ | 0.262 |
| No | 418 (25.68) | 387 (26.20) | 31 (20.53) |  |  |
| Yes | 544 (33.41) | 487 (32.97) | 57 (37.75) |  |  |
| Unknown | 666 (40.91) | 603 (40.83) | 63 (41.72) |  |  |
| Smoking |  |  |  | － | 0.020|| |
| Never | 848 (52.09) | 754 (51.05) | 94 (62.25) |  |  |
| Current | 543 (33.36) | 497 (33.65) | 46 (30.46) |  |  |
| Ever | 220 (13.51) | 210 (14.22) | 10 (6.62) |  |  |
| Unknown | 17 (1.04) | 16 (1.08) | 1 (0.66) |  |  |
| Drinking |  |  |  | 0.71§ | 0.870 |
| Every day | 415 (25.49) | 380 (25.73) | 35 (23.18) |  |  |
| 3–4 times/week | 107 (6.57) | 98 (6.64) | 9 (5.96) |  |  |
| ≤1–2 times/week | 248 (15.24) | 225 (15.23) | 23 (15.23) |  |  |
| Unknown | 858 (52.70) | 774 (52.40) | 84 (55.63) |  |  |
| Number of types of chronic diseases\* |  |  |  | － | 0.132|| |
| 0 | 1002 (61.55) | 902 (61.07) | 100 (66.23) |  |  |
| 1 | 473 (29.05) | 432 (29.25) | 41 (27.15) |  |  |
| 2 | 130 (7.99) | 123 (8.33) | 7 (4.64) |  |  |
| 3 | 21 (1.29) | 19 (1.29) | 2 (1.32) |  |  |
| 4 | 2 (0.12) | 1 (0.07) | 1 (0.66) |  |  |
| 5 | 0 (0) | 0 (0) | 0 (0) |  |  |

The values are presented as means ±  standard deviations median (Quarter 1–Quarter 3), or counts (%). \*Chronic diseases included diabetes, hypertension, chronic obstructive pulmonary disease, coronary heart disease, and stroke. BMI: Body mass index.

†*t* values, ‡*U* values, §*χ2* values, || Fisher’s exact test

**Supplementary Table 2: Association between sleep duration and testosterone level among different BMI groups according to adjusted multivariate linear regression models.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | BMI <25 |  | 25 ≤ BMI < 30 |  | BMI ≥30 |  |
| Items | **β (95% CI)** | ***P***-**value** | **β (95% CI)** | ***P***-**value** | **β (95% CI)** | ***P-*value** |
| Sleep duration\* | 134.98 (7.90–262.06)  | 0.038 | 3.79 (−9.76–17.34)  | 0.584 | −16.80 (−67.62–34.03)  | 0.525 |
| *P* for interaction | 0.010 |  |

Adjusted for age, ethnic group, education level, marital status, smoking history, alcohol use, nutrition status, and physical frailty. BMI: Body mass index; CI: Confidence interval. \*Change in testosterone for each additional hour of sleep duration.

***Results of sensitivity analyses***

1. In the imputed population, the regression analysis results were consistent with the results in main text. A longer sleep duration was still significantly associated with increased TT concentrations (612.72 ng/dL, 95% Confidence Interval (CI): 234.59–990.85, P=0.002). [Supplementary Table 3]

**Supplementary Table 3: After imputing missing values, the association between sleep duration and TT level according to adjusted multivariate linear regression models.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sleep duration (h) | Model 1 |  | Model 2 |  |
| **β (95% CI)** | ***P***-**value** | **β (95% CI)** | ***P***-**value** |
| <9.5 | Ref. | － | Ref. | － |
| ≥9.5 | 409.98 (161.81–658.14)  | 0.001 | 612.72 (234.59–990.85)  | 0.002 |

Model 1: Adjusted for BMI, age, ethnic group, education level, marital status, smoking history, and alcohol use. Model 2: Adjusted for BMI, age, ethnic group, education level, marital status, smoking history, alcohol use, cognitive performances, nutritional status, physical frailty, and number of chronic diseases. BMI: Body mass index; CI: Confidence interval; TT: Total testosterone.

1. Using a cut-off value of 10 PSQI scores, we divided participants into groups of the normal sleep quality group and the poor sleep quality group. We found that sleep quality was not independently associated with TT concentrations (39.08, 95% CI: –528.30–606.46, P=0.892). [Supplementary Table 4]

**Supplementary Table 4: Association between sleep quality and the levels of TT according to adjusted multivariate linear regression models.**

|  |  |  |
| --- | --- | --- |
| Sleep quality | Model 1 | Model 2 |
| **β (95% CI)** | ***P***-**value** | **β (95% CI)** | ***P***-**value** |
| Normal  | Ref. | － | Ref. | － |
| Poor | 68.75 (−302.90−440.39)  | 0.717 | 39.08 (−528.30−606.46)  | 0.892 |
| PSQI score | −17.05(−48.42−14.31)  | 0.287 | −23.52(−71.32−24.29)  | 0.336 |

Model 1: Adjusted for BMI, age, ethnic group, education level, marital status, smoking history, and alcohol use. Model 2: Adjusted for BMI, age, ethnic group, education level, marital status, smoking history, alcohol use, cognitive performances, nutritional status, physical frailty, and number of chronic diseases. BMI: Body mass index; CI: Confidence interval; PSQI: Pittsburgh Sleep Quality Index; TT: Total testosterone. Normal: PSQI <10; poor: PSQI ≥10.

1. In a multivariable logistic regression analysis, we examined the relationship between the sleep duration and LOH. The sleep duration was not independently significantly associated with the prevalence of TT concentrations <300 ng/dL after fully adjusting for covariables (Odd ratio (OR): 1.12 [0.47–2.69], P=0.80) [Supplementary Table 5].

**Supplementary Table 5: Association between sleep duration and the prevalence of LOH according to adjusted multivariate logistic regression models.**

|  |  |  |
| --- | --- | --- |
| Sleep duration (h) | Model 1 | Model 2 |
| **OR (95% CI)** | ***P***-**value** | **OR (95% CI)** | ***P***-**value** |
| <9.5 | Ref. | － | Ref. | － |
| ≥9.5 | 1.14 (0.58–2.24)  | 0.698 | 1.12 (0.47–2.69)  | 0.801 |
| Sleep duration\* | 1.03 (0.92–1.15)  | 0.646 | 0.96 (0.83–1.12)  | 0.624 |

Model 1: Adjusted for BMI, age, ethnic group, education level, marital status, smoking history, and alcohol use. Model 2: Adjusted for BMI, age, ethnic group, education level, marital status, smoking history, alcohol use, cognitive performances, nutritional status, physical frailty, and number of chronic diseases. BMI: Body mass index; CI: Confidence interval; LOH: Late-onset hypogonadism TT: Total testosterone. \* Change in testosterone for each additional hour of sleep duration.

1. Because the above-mentioned analyses determined 9.5 h as the cut-off value of long and short sleep, we selected 7, 8, and 9 h as alternative cut-off values for a sensitivity analysis. Multivariable linear regression analyses showed that when 7, 8, and 9 h were selected as cut-off values, there was no significant association between the sleep duration and TT concentrations. [Supplementary Table 6]

**Supplementary Table 6: Association between different sleep durations and TT level according to adjusted multivariate linear regression models.**

|  |  |  |
| --- | --- | --- |
| Sleep duration (h) | Model 1 | Model 2 |
| **β (95% CI)** | ***P***-**value** | **β (95% CI)** | ***P***-**value** |
| <9.5 | Ref. | － | Ref. | － |
| ≥9.5 | 436.48 (181.81–691.15)  | <0.001 | 617.84 (227.45–1008.23)  | 0.002 |
| <9 | Ref. | － | Ref. | － |
| ≥9 | 155.65 (−13.85, 325.14) | 0.072 | 233.84 (−29.34, 497.02) | 0.082 |
| <8 | Ref. | － | Ref. | － |
| ≥8 | 66.11 (−78.04, 210.26) | 0.369 | 120.87 (−93.77, 335.52) | 0.270 |
| <7 | Ref. | － | Ref. | － |
| ≥7 | 59.98 (−112.52, 232.48) | 0.495 | 65.62 (−194.31, 325.56) | 0.620 |

Model 1: Adjusted for BMI, age, ethnic group, education level, marital status, smoking history, and alcohol use. Model 2: Adjusted for BMI, age, ethnic group, education level, marital status, smoking history, alcohol use, cognitive performances, nutritional status, physical frailty, and number of chronic diseases. BMI: Body mass index; CI: Confidence interval; TT: Total testosterone.

1. After matching, we obtained a new cohort of 1:3 between the long sleep and the short sleep groups. There were no significant differences in baseline characteristics between the two groups in the matched cohort, except for alcohol and tobacco use and education. The long sleep group still had higher TT concentrations than the short sleep group. The results of multivariable linear regression were also consistent with pre-matching, and the long sleep group was significantly associated with higher TT concentrations than the short sleep group. [Supplementary Table 7 and 8]

**Supplementary Table 7: Baseline characteristics of participants with different sleep duration after propensity score matching.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Characteristics | Total(*n*=567) | Sleep duration | *Statistics* | *P*-value |
| **< 9.5 h****(n=425)** | **≥9.5 h****(n=142)** |
| Age (years) | 68.16 ± 6.22 | 67.87 ± 6.03 | 69.04 ± 6.69 | –1.94† | 0.053 |
| BMI (kg/m2) | 24.67 ± 4.04 | 24.73 ± 4.13 | 24.49 ± 3.75 | 0.62† | 0.537 |
| Testosterone (ng/dL) | 434.00(313.00－568.90) | 425.30 (298.00－560.00) | 444.50 (324.50－580.00) | –1.23‡ | 0.22 |
| Education |  |  |  | 12.80§ | 0.012 |
| No | 98 (17.28) | 62 (14.59) | 36 (25.35) |  |  |
| Primary school | 219 (38.63) | 171 (40.24) | 48 (33.80) |  |  |
| Junior high school | 132 (23.28) | 102 (24.00) | 30 (21.13) |  |  |
| High school | 88 (15.52) | 63 (14.82) | 25 (17.61) |  |  |
| Undergraduate and above | 30 (5.29) | 27 (6.35) | 3 (2.11) |  |  |
| Marital status |  |  |  | － | 0.688|| |
| Unmarried | 4 (0.71) | 3 (0.71) | 1 (0.70) |  |  |
| Married | 511 (90.11) | 385 (90.58) | 126 (88.74) |  |  |
| Divorce | 4 (0.71) | 2 (0.47) | 2 (1.41) |  |  |
| Widowed | 48 (8.47) | 35 (8.24) | 13 (9.15) |  |  |
| Smoking |  |  |  | 104.26§ | <0.001 |
| No | 179 (31.57) | 89 (20.94) | 90 (63.38) |  |  |
| Yes | 203 (35.80) | 159 (37.41) | 44 (30.99) |  |  |
| Used to | 185 (32.63) | 177 (41.65) | 8 (5.63) |  |  |
| Drinking |  |  |  | 13.31§ | 0.001 |
| Every day | 242 (42.68) | 163 (38.35) | 79 (55.64) |  |  |
| 3–4 times/week | 113 (19.93) | 89 (20.94) | 24 (16.90) |  |  |
| ≤1–2 times/week | 212 (37.39) | 173 (40.71) | 39 (27.46) |  |  |
| Number of chronic diseases, *n*\* |  |  |  | － | 0.075|| |
| 0 | 355 (62.60) | 259 (60.94) | 96 (67.61) |  |  |
| 1 | 158 (27.87) | 120 (28.24) | 38 (26.76) |  |  |
| 2 | 45 (7.94) | 40 (9.41) | 5 (3.52) |  |  |
| 3 | 8 (1.41) | 6 (1.41) | 2 (1.41) |  |  |
| 4 | 1 (0.18) | 0 (0.00) | 1 (0.70) |  |  |
| 5 | 0 (0) | 0 (0) | 1. (0)
 |  |  |

The values are presented as means ± SDs, median (Quartile1- Quartile3) or counts (%). BMI: Body mass index. \*Chronic diseases included diabetes, hypertension, chronic obstructive pulmonary disease, coronary heart disease, and stroke.

†*t* values, ‡*U* values, §*χ2* values, || Fisher’s exact test

**Supplementary Table 8: In matched population, the association between sleep duration and TT level according to adjusted multivariate linear regression models.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sleep duration (h) | Model 1 |  | Model 2 |  |
| **β (95% CI)** | ***P***-**value** | **β (95% CI)** | ***P***-**value** |
| <9.5 | Ref. | － | Ref. | － |
| ≥9.5 | 316.91 (74.03–559.78)  | 0.011 | 447.40 (46–848.80)  | 0.030 |

Model 1: Adjusted for BMI, age, ethnic group, education level, marital status, smoking history, and alcohol use. Model 2: Adjusted for BMI, age, ethnic group, education level, marital status, smoking history, alcohol use, cognitive performances, nutritional status, physical frailty, and number of chronic diseases. BMI: Body mass index; CI: Confidence interval; TT: Total testosterone.

***Results of power of test***

When the mean level (TT levels of long sleep duration group) among the general population was 663.13 ng/dL, the standard deviation was 2201.36 ng/dL, the prevalence of exposure among the general population was 0.1, and the *post hoc* statistics power was 0.924.

**References**

1. Vesper HW, Botelho JC. Standardization of testosterone measurements in humans. J Steroid Biochem Mol Biol 2010;121:513-519. doi: 10.1016/j.jsbmb.2010.03.032.

2. Zhang C, Bian H, Chen Z, Tian B, Wang H, Tu X, *et al*. The association between dietary inflammatory index and sex hormones among men in the United States. J Urol 2021;206:97–103. doi: 10.1097/JU.0000000000001703.

3. Huang Y, Zhu M. Increased Global PSQI Score is associated with depressive symptoms in an adult population from the United States. Nat Sci Sleep 2020;12:487–495. doi: 10.2147/NSS.S256625.

4. Marks LS, Mazer NA, Mostaghel E, Hess DL, Dorey FJ, Epstein JI, *et al*. Effect of testosterone replacement therapy on prostate tissue in men with late-onset hypogonadism: A randomized controlled trial. JAMA. 2006;296:2351–2361. doi: 10.1001/jama.296.19.2351.