# Evaluating Calculated Free Testosterone as a Predictor of Morbidity and Mortality Independent of Testosterone for Cross-sectional and 5 year Longitudinal Health Outcomes in Older Men: The Concord Health and Ageing in Men Project

Benjumin Hsu[1,2,3], Robert G Cumming[1,2,3,4], Fiona M Blyth[2], Vasi Naganathan[2], Louise M Waite[2], David G Le Couteur[1,2] , Markus J Seibel[1], David J Handelsman[1]

[1] ANZAC Research Institute, University of Sydney and Concord Hospital, Sydney, New South Wales, Australia

[2] Centre of Education and Research on Ageing, University of Sydney and Ageing and Alzheimer's Institute, Concord Hospital, Sydney, New South Wales, Australia

[3] ARC Centre of Excellence in Population Ageing Research, University of Sydney, Sydney, New South Wales, Australia.

[4] School of Public Health, University of Sydney, Sydney, New South Wales, Australia

**Brief Title:** Evaluating Calculated Free Testosterone

**Address all correspondence and requests for reprints to:** David J. Handelsman, MB BS PhD, ANZAC Research Institute, Sydney, New South Wales, Australia 2139. E-mail: djh@anzac.edu.au Phone: +61 2 9767 9100

**Funding:** The CHAMP study is funded by the NHMRC Project Grant (No. 301916), Sydney Medical School Foundation and Ageing and Alzheimer's Institute.

**Disclosure:** D.G.L.C. is a co-deputy editor for the Journal of Gerontology: Biological Sciences. All the other authors have nothing to declare.

**ABSTRACT**

To determine whether calculated free testosterone (cFT) provides prognostic information independent of serum T for predicting morbidity and mortality in older men in cross-sectional and 5-year longitudinal analyses. We studied men aged ≥70 years at baseline (n=1705), 2-year and 5-year measuring serum T (liquid chromatography-mass spectrometry), SHBG (immunoassay), cFT (an assumption-free empirical formula) together with 24 morbidity and 4 mortality outcomes. For cross-sectional and longitudinal analyses we employed a joint prediction model using generalized estimating equation models adjusted for age, smoking, comorbidities and BMI with men having both normal T and normal cFT as referent group. Most morbidity and mortality outcomes were predicted by a combination of low T and cFT (LL). By contrast, only a single morbidity outcome in cross-sectional and none in longitudinal analysis was predicted by low T/normal cFT (LN) or normal T/low cFT (NL) without significant LL associations (isolated discordance). While for the few outcomes that predicted morbidity in men with discordances (LN or NL), these predictions only occurred when LL was also significant. Hence, for morbidity or mortality prediction in older men, discordance between cFT and T is unusual and isolated discordance is rare, so that cFT provides minimal independent prognostic information over serum T.

**Keywords:** reproductive hormones, androgen, epidemiology, health outcomes, signs and symptoms

## INTRODUCTION

In men, most circulating testosterone (T) is bound to SHBG with the remainder bound to albumin and other low-affinity binding proteins and only 1-2% unbound to any circulating protein. The Free Hormone Hypothesis (FHH) postulates that this small unbound moiety is the most biologically active fraction of circulating serum T for its greater accessibility to tissues(1). Yet this theory cannot explain why unbound hormones are more rather than less biologically active as they are also more accessible to sites of degradation than bound hormones. Yet, despite its wide adoption, the FHH remains unproven and almost untested (2). FHH might have an empirical basis if FT provides additional independent biological or clinical information independent of serum T measurement for androgen responsive health outcomes. There however has not been a systematic empirical evaluation of free testosterone (FT) measurement, for example, in predicting morbidity or mortality outcomes independent of accurate T measurement by mass spectrometry (MS)-based methods (3, 4).

As dialysis-based laboratory measurement of FT is a laborious and exacting manual method, it is rarely available so that various formulae are substituted to calculate FT (cFT) (5-7). However, comparative evaluations based on laboratory FT measurement as the gold standard show that the widely used model-based formulae (Sodergard, Vermeulen) are inaccurate due to their obligatory assumptions of plug-in estimates for the stoichiometry and affinity of testosterone binding to SHBG (5-10). We therefore developed and extensively validated an assumption-free, fully empirical formula for cFT (FTZ) that does not require plug-in estimates of binding stoichiometry

and affinity of testosterone for SHBG (5-7). Therefore the present study has primarily used this formula with comparison against a more widely used model-based (Vermeulen) formula (11).

In multiple studies of older men in the Concord Health and Ageing in Men Project (CHAMP) (12-19) and Health in Men Study (HIMS) (20-25) cohorts, effect size and association of health outcomes based on accurate cFT estimates using the FTZ formula appeared not to diverge substantially from those based on serum testosterone measurement by LC-MS. Hence in this study, we aimed to determine formally whether accurately calculated FT provides additional prognostic information independent of serum T measured by LC-MS in predicting morbidity and mortality in older men in both cross-sectional and 5 year longitudinal analyses. As cFT is a deterministic function of T by its formula, we utilized a pattern of joint prediction to evaluate independent predictive contributions of health outcomes while avoiding collinearity.

## METHODS

### Study Participants

The CHAMP study is a longitudinal, population-based observational study of male ageing conducted among men living in the vicinity of Concord Hospital in Sydney, New South Wales, Australia as described in detail previously (26). Community dwelling men aged at least 70 years in 2005 were eligible with no other inclusion or exclusion criteria resulting in a final inception cohort of 1705 participants. Baseline measurements

were conducted between January 2005 and June 2007 using self-reported and interviewer-administered questionnaires and a wide range of clinical assessments. Follow-up assessments were conducted between January 2007 and October 2009 for 2-year follow-up, and August 2010 and July 2013 for the 5-year follow-up, with identical measurements as at baseline. All participants gave written informed consent. The study was approved by the Sydney South West Area Health Service Human Research Ethics Committee, Concord Repatriation General Hospital, Sydney, Australia.

**Reproductive Hormone Measurement**

Participants had an early morning fasting blood sample taken at baseline with serum stored at -80 C until assay. Measurement of serum T was by liquid chromatography-tandem mass spectrometry (LC-MS) as described (27) with modifications by introducing ultrapressure for high pressure liquid chromatography with corresponding changes in extraction methodology validated according to FDA criteria (for details see supplementary methods in (28)). The steroid measurements were calibrated against certified reference materials for T (National Measurement Institute, North Ryde, Australia). The assays had between-run coefficients of variation (CV) at three levels (low, medium, high) of quality control (QC) specimens of 1.9-4.5%, 3.8-7.6%, 2.9-13.6% and 5.7-8.7%, respectively, over 224 runs including all samples from this study. Overlapping QC samples were routinely run at the start, middle and end of every run with each new QC control run multiple times for calibration before use and

there was no evidence of assay drift (13). Serum SHBG were measured by automated immunoassays (Roche Diagnostics Australia, Dee Why, Australia) subject to ongoing external QC program calibration with between-assay CV for 2 levels of QC specimens in each run of 2.0-2.8% for SHBG. The cFT levels in this study were computed using an assumption-free, empirical formula (FTZ) developed and validated against laboratory-based measurements of FT by dialysis methods which have displayed much closer conformance with laboratory-measured FT than model-based formulae (6, 7).

**Morbidity and Mortality Outcomes Measurement**

Health-related quality of life and self-rated health were assessed using the 12-Item Short Form Health Survey (SF-12) (29). Functional disability was defined using the Katz activity of daily living (ADL) questionnaire (30). Frailty was defined according to the criteria used in the CHS: weight loss/shrinking, weakness, exhaustion, slowness and low activity (31). Falls were measured at the four month follow-up phone calls after their baseline assessments, participants were asked whether they had fallen in the preceding 4 months and, if so, how many times they had fallen. Participants were assessed for cognitive impairment at the clinic assessment visits using the Mini Mental State Examination (MMSE) (32). Depressive symptoms were evaluated by the Geriatric Depression Scale (GDS), short form (33). The participants were asked about erectile dysfunction, sexual activity, sexual desire and sexual satisfaction using standard, validated questionnaires (34). Metabolic syndrome was defined using the NCEP Adult

Treatment Panel (ATP) III criteria (35). Physical activity was measured using the Physical Activity Scale for the Elderly (PASE) (36). Walking speed was measured at the participants' usual pace (31). Trained staff used a stopwatch to record the time taken by the men to walk 6 meters. The fastest time from two trials was used. Bone mineral density (BMD) at the total hip and femoral neck, lean mass and body fat was measured using dual X-ray absorptiometry (DXA; Discovery-W scanner Hologic, Bedford, MA). The appendicular lean mass (ALM) was calculated as the sum of lean mass of arms and legs (kg) (37). The ALM was standardized by BMI ($ALM_{BMI}$) to take into account the body size of participants (38). Handgrip strength was measured with a Jamar dynamometer (Promedics, Blackburn, United Kingdom). Weight (by a regularly calibrated scale), height (using a Harpenden stadiometer) and waist circumference were measured by a trained professional at the clinic visit. Fasting blood samples were obtained at each visit for biochemistry tests including hemoglobin, glucose and PSA performed at the accredited Clinical Pathology department of Concord Hospital. The New South Wales Registry of Births, Deaths, and Marriages was contacted to ascertain death status. The Registry also provided details recorded on the original death certificate for all participants. Based on the information provided from the death certificates, the general underlying cause of death (cancer, cardiovascular or other) was identified independently by two medical practitioners (RGC, DJH) (39).

**Potential Confounder Measurement**

Tobacco usage status (current, ex- or never smoker) was by self-reported questionnaires. A comorbidity score was calculated as the sum of all conditions reported from the 19 disorders listed in the questionnaire. Body mass index (BMI) was calculated from clinic measurements of height and weight.

**Statistical Analysis**

Descriptive characteristics of reproductive hormones at baseline and study health outcomes at baseline, 2-years and 5-years follow-up were generated for the analytic sample (table 1). Participants were categorized into four mutually exclusive groups based on their baseline serum T and cFT defining "low" for these analyses by setting a threshold of the lowest quintile ($20^{th}$ centile) for serum T (10.2 nM) and cFT (156 pM). The referent group for all cross-sectional and longitudinal analyses were men with both normal serum T and cFT (NN) with the other groups defined as men with the combinations of normal T/low cFT (NL), low T/normal cFT (LN) or low T/low cFT (LL). Of the 1705 participants who completed the baseline assessments, a total of 1651 were included for analyses in this paper, after excluding men using androgen or anti-androgen treatments (n=20) or with missing data on reproductive hormones (n=34).

Joint prediction in cross-sectional associations between the T/cFT status and health outcomes were assessed by logistic regression for categorical health outcomes, by multiple regression for continuous health outcomes and by Cox regression for mortality outcomes. Results were summarized into concordant findings (only

statistically significant LL), discordant findings (either statistically significant LN and LL, or statistically significant NL and LL), and isolated discordance findings (statistically significant LN or NL without statistically significant LL). The detailed results for categorical variables are presented as odds ratios (95% confidence interval), for continuous variables are presented as β-values (95% confidence interval) and for mortality are presented as hazard ratios (95% confidence interval).

Similarly, longitudinal association between baseline T/cFT status and changes in health outcomes across baseline, 2-years and 5-years were assessed by generalized estimating equations (GEE) with exchangeable working correlation and robust variance estimator. The multinomial cumulative logit model was used for categorical morbidity outcomes, linear model for continuous morbidity outcomes and Poisson loglinear model for mortality outcomes. GEE method is robust and efficient when treating missing data in longitudinal studies (40).

Model building for both cross-sectional and longitudinal analyses included adjustment for known major covariates, notably, age, body mass index (BMI), smoking status and comorbidity. BMI was not adjusted for in analyses for metabolic syndrome, body fat, weight and waist circumference analysis. For post-hoc analyses, a Bonferroni adjustment to notional p-values was performed to account for multiple comparisons involved in evaluating 28 outcome comparisons from a single set of data so that the conventional 0.05 level of significance was adjusted to a threshold of 0.002 (0.05/28). Models were fitted using SPSS software version 20 (IBM Corp., Armonk, NY, USA) and SAS software 9.3 (SAS Institute Inc., Cary, NC, USA).

## RESULTS

The demographic and anthropometric details of the CHAMP cohort are provided in Table 1 and the descriptive details of the T, SHBG and cFT in Table 2. A total of 1283 men (78%) were categorized into normal T/normal cFT (NN), 40 men (2%) into normal T/low cFT (NL), 38 men (2%) into low T/normal cFT (LN), and 290 men (18%) into low T/low cFT (LL).

### Cross-sectional morbidity analysis

The baseline cross-sectional associations between T/cFT status and morbidity outcomes are shown in detail in supplementary table 1 and summarized in table 3 and figure 1. After multivariable adjustment of the baseline cross-sectional data, low T/cFT (LL) was significantly associated with 15 of 24 outcomes (frailty, falls, sexual satisfaction, sexual desire, sexual activity, metabolic syndrome, physical activity, walking speed, physical quality of life, weight, hip BMD, body percent fat, waist circumference, glucose, hemoglobin and PSA). Where LL was not a significantly associated with outcomes, there was only a single morbidity outcome that remained associated with either LN or NL (isolated discordance). A few morbidity outcomes displayed significant associations with discordant findings - 4 outcomes (metabolic syndrome, weight, fat mass and waist circumference) for LN and 3 outcomes (weight, fat mass and waist circumference) for NL – but for each of these outcomes LL was also significant. After Bonferroni correction, LL

remained significantly associated with 8 of 24 outcomes in cross-sectional analysis with additional discordant findings in 4 outcomes for LN and no outcomes for NL. There were no associations with isolated discordance.

**Longitudinal morbidity analysis**

The longitudinal associations over the 5-year follow-up between baseline T/cFT status and changes in morbidity outcomes are shown in detail in supplementary table 2 and summarized in Table 3 and figure 1. Similar to the cross-sectional analysis, in multivariate adjusted models of the 5 year longitudinal analysis, low T/cFT (LL) was statistically significantly associated with 16 of 24 outcomes (poor self-rated health, ADL disability, frailty, sexual satisfaction, sexual desire, sexual activity, metabolic syndrome, physical activity, walking speed, hip BMD, physical quality of life, weight, body percent fat, waist circumference, glucose and hemoglobin). Significant discordant findings were present in 6 outcomes (metabolic syndrome, weight, fat mass, waist circumference, glucose and hemoglobin) for LN and 4 (sexual activity, hip BMD, fat mass and hemoglobin) for NL where LL was also significant. There were no significant isolated discordant findings (significant LN or NL without significant LL). After Bonferroni correction, LL remained significantly associated with 9 of 24 morbidity outcomes with additional discordant findings in 4 outcomes for LN and no outcomes for NL, all in conjunction with significant LL.

**Sensitivity analysis**

One sensitivity analysis performed used the same empirical formula (FTZ) but lowering the threshold to define "low" from lowest quintiles (lowest 20%) to lowest centiles (lowest 10%) for the morbidity analysis produced essentially the same results. In the multivariable adjusted model, LL was significantly associated with 12 of 24 morbidity outcomes in cross-sectional and 15 of 24 outcomes in longitudinal analysis (table 3). Additional significant discordant findings (LN or NL) in conjunction with significant LL were present for 6 outcomes in cross-sectional and 15 outcomes in longitudinal analysis. When LL was not significant (isolated discordance), only 1 outcome in cross-sectional and 1 in longitudinal analysis were significantly associated with an adverse morbidity outcome.

Another sensitivity analysis was conducted using the Vermeulen cFT (lowest quintile) for the same morbidity analysis and the findings were similar to our original analysis using the more accurate FTZ formula. In multivariable model, LL was significantly associated with 12 of 24 morbidity outcomes in cross-sectional and 15 of 24 outcomes in longitudinal analysis. Where LL was not a significant predictor, isolated discordant findings (either LN or NL significant) were associated with only 1 outcome in cross-sectional and 1 in longitudinal analyses. Among men with discordant findings (significant LN or NL, with significant LL), morbidity prediction was present for 6 outcomes in cross-sectional and 15 outcomes in longitudinal analysis.

**Cross-sectional and longitudinal mortality analysis**

With multi-variable adjustment, both the baseline cross-sectional as well as the longitudinal mortality analyses (supplementary table 3) showed significant effects for LL in all-cause, cardiovascular and other but not for cancer mortality. For no mortality outcome was there significant prediction when LL was not significant and where mortality outcomes were predicted by discordant (NL or LN), LL was also significant.

## DISCUSSION

Many studies have reported associations between health outcomes and low T or low cFT, considered as separate parameters, among older men (41). However, we observed consistently in a series of studies from the CHAMP (12-19) and HIMS (20-25) cohorts that cFT as a predictor rarely, if ever, provided any significantly different information on health outcomes from serum T measured by LC-MS. Furthermore, as the FHH remains largely untested, there remains minimal critical evidence to what extent, if any, FT data provides additional biological or clinical insight independent of accurate serum T measurements by LC-MS in men (42). The present findings investigating a wide range of morbidity and mortality outcomes in older men suggest that cFT rarely adds independent prognostic information to serum T measured by LC-MS in either cross-sectional or longitudinal analyses.

An important caveat is that the utility of health outcome predictions by T and/or FT, depends on the accuracy of the T and FT estimates employed. Until recently, most

studies relied upon T immunoassays which suffer from method-specific and other technical limitations notably if applied to reduced serum testosterone such as in older men (3, 4). This became a greater problem over the decades after the 1980's when direct, non-extraction immunoassay became almost universal in clinical practice and research. Over the last decade, more accurate measurement of serum T has become more widely feasible using modern, bench-top LC-MS to supplant direct (unextracted) T immunoassays.

Currently, especially for large scale epidemiological studies, FT is rarely measured directly by dialysis-based laboratory reference methods. These methods are laborious, exacting and require manual laboratory skills which have been largely eliminated by the deskilling automation of chemical pathology laboratories. Furthermore FT measurements lack quality control programs or validated reference ranges. Instead, FT is usually calculated by a variety of formulae which fall into two classes, model-based equilibrium binding and fully empirical equations. These differ in their assumptions and in conformance in accuracy to dialysis-based laboratory gold standard reference methods. The accuracy of cFT is crucial because any formula produces a deterministic (inverse) function of age as it compounds two age-dependent variables – testosterone and SHBG. Unless the formula accurately represents the authentic laboratory-based FT measurement it intends to represent, it will display a spurious correlation with any age-dependent variable regardless of whether that variable has any genuine biological relationship to testosterone.

312 The FTZ equation was originally developed from a large dataset of 3975 serum
313 samples by identifying the optimal regression formula of laboratory dialysis-based
314 reference FT measurements on serum testosterone and SHBG measured in the same
315 samples. This formula was cross-validated against a separate set of 124 serum samples
316 (6) and then subsequently confirmed as highly accurate when tested in a different large
317 dataset of 2159 samples from another laboratory using different methods to measure
318 FT, testosterone and SHBG (7). A key finding from the extensive validation involving over
319 6000 serum samples was that the widely used, model-based equilibrium binding
320 equation-based formulae by Vermeulen (11) and Sodergard (43) display marked bias
321 deviating from the laboratory-measured FT. These deviations were due to both wrong
322 stoichiometry as well as arbitrary plug-in binding affinity coefficients for T binding to
323 SHBG (6, 7), the latter varying 5-fold among the various implementations of model-
324 based equilibrium binding formulae (44).

325 The present study uses this FTZ formula to evaluate the impact of accurately
326 estimated cFT, corresponding most closely to laboratory-based FT measurements, on
327 morbidity and mortality outcome predictions. The novelty of the current longitudinal
328 study is that it investigates both serum T and cFT levels concurrently as joint predictors
329 of a wide range of health outcomes over time. Our analysis revealed that both cross-
330 sectionally and longitudinally over 5 years, men with concordant low serum T and cFT as
331 well as those with concordant normal T and cFT were more likely to die or experience
332 adverse health outcomes. On the contrary, only a minority of men had variables
333 displaying discrepancies between T and cFT values and where there was an isolated

discordance – that is discordance between T and cFT but not accompanied significant association or prediction by LL - was rare. Hence, not only are discrepancies unusual but cFT alone predicts almost no health outcomes among older men independent of an accurately measured serum T. Altogether, the present analysis provides a comprehensive analysis of a wide range of health outcomes including non-specific symptoms resembling those of androgen deficiency or many other chronic diseases.

A recent study from the EMAS cohort evaluated the FHH among older men by analyzing cross-sectionally the joint association of cFT and T with health outcomes. They reported that low Vermeulen cFT, even in the presence of normal T, but not the combination of normal cFT and low T, was associated with a range of non-specific symptoms including sexual and physical symptoms (45). In a previous study they reported that low T and cFT were associated with sexual but not physical or psychological symptoms (46) although the sexual symptoms had high rates of false positive and false negatives reflecting their non-specificity and the direction of causality could not be determined. This reflects the fact that genuine androgen deficiency symptoms are, for any individual, are highly reproducible at consistent blood testosterone concentrations (47) ; however, as the actual symptoms differ widely between individuals, grouping individuals according to symptoms erodes the relationship of symptoms to blood testosterone concentrations (48). Furthermore, as the Vermeulen model-based cFT formula systematically deviates from laboratory measured FT values as reported by several independent groups (5-10). Yet, as any cFT remains a deterministic (inverse) function of age, failure to correspond accurately to

laboratory-measured FT makes it likely that any relationship to age-related symptoms may reflect residual confounding due to the age-mismatch of the subgroups (persisting after linear age adjustment) rather than any authentic relationship with serum T. The present analysis, using a more accurate and extensively validated cFT formula so that it corresponds more closely to laboratory-measured FT, showed that cFT and T were usually concordant and, in the unusual instances where there was a discordance, that almost always occurred only when the concordant low T/low cFT was also significant. In our analysis as well as that using the Vermeulen formula, significant isolated discordant association or prediction by a low cFT was rare and had little impact on prediction of mortality or morbidity over the next 5 years. Instead it was the combination of both a low T and low cFT that was significantly associated with most outcomes although the direction of causality remains undetermined.

Our study shows that men with low T were most likely to have low cFT while men with normal T were most likely to have normal cFT. Only a very small proportion had discordance with either normal T and low cFT, or low T and normal cFT and when this occurred it was almost invariably in the setting where the concordant combination of low T/low CFT was also significant. The major finding in this study is consistent with previous studies showing low T and low cFT, as separate parameters, are associated with these many health outcomes such as general health status, functional ability, metabolic syndrome, bone health, cognition, sexual function, etc. These findings confirm our impression from previous studies in the CHAMP (12-19) and HIMS (20-25) cohorts that show very similar effect size and associations in either low T or low cFT with

a wide variety of health outcomes (12-25). This suggest that cFT provides minimal independent predictive information for health outcomes independent of accurately measured serum T and questions whether cFT estimates, even when accurately calculated, provide any useful information for clinical practice.

The strengths of this study include the use of longitudinal data to investigate a comprehensive profile of T/cFT status in conjunction with a wide array of key morbidity and mortality outcomes over three follow-up time-points spanning 5-years. Another is the use of the LC-MS, the current gold standard for steroid assays, providing multi-analyte steroid profiling. This improves upon direct immunoassay methods which, lacking extraction and chromatography, feature poor accuracy at low levels of circulating sex steroids, which is particularly problematic for measuring circulating T in older men (3, 4). Furthermore, we used an extensively validated, assumption-free formula for cFT which corresponds more accurately to laboratory-measured FT than previous model-based equilibrium binding formula that rely on arbitrary plug-in coefficients. A further strength of CHAMP is that it includes a large and representative group of older Australian men, as demonstrated by similar socio-demographic and health characteristics compared to older men in the nationally representative MATeS study (49).

A significant limitation of our study is the impact of survivor bias. This applies to the survivorship in the cohort with most losses due to mortality which accounted for nearly 35% of loss to follow-up in our cohort. On the other hand, mortality was evaluated as an outcome so that this cohort provides a more complete view of the

causes and determinant of mortality among living older men. To avoid the impact of potential diurnal variation in hormone concentrations, a rhythm that is mostly lost in ageing men (50), we obtained fasting morning blood samples in this study and evaluated joint prediction to avoid collinearity between cFT and T.

In conclusion, concordant low serum T and cFT levels were strongly associated with many health outcomes in older men whereas among the minority of men with discrepancies between T and cFT, such discordance was associated with or predicted few health outcomes and only then when for the same outcome, there was also a significant association or prediction by the combination of both low T and cFT. Hence, in addition to the ambiguous theoretical basis of the FHH, the present findings suggest that even accurately cFT estimates provide minimal additional clinical or biological information independent of accurate measurement of serum T concentrations for mortality or morbidity outcomes in older men. These findings provide little support for the application of the FHH to studies of testosterone and clinical outcomes in older men.

**ACKNOWLEDGMENTS**

R.G.C., D.J.H., M.J.S., L.M.W., V.N., D.G.L.C. and F.M.B. contributed to the formulation of the study concept, design, methods, subject recruitment and data collection. B.H. wrote the manuscript and performed the analyses. D.J.H. wrote portions

of the manuscript. R.G.C., F.M.B., V.N., D.G.L.C., M.J.S., and L.M.W. reviewed the manuscript and contributed to discussion.

The CHAMP study is funded by the NHMRC Project Grant (No. 301916), Sydney Medical School Foundation and Ageing and Alzheimer's Institute.

## REFERENCES

1. Goldman AL, Bhasin S, Wu FCW, Krishna M, Matsumoto AM, Jasuja R. A Reappraisal of Testosterone's Binding in Circulation: Physiological and Clinical Implications. Endocr Rev. 2017;doi.10.1210/er.2017-00025
2. Handelsman D. Free Testosterone: Pumping up the Tires or Ending the Free Ride? Endocr Rev. 2017;**38**:297-301;doi.10.1210/er.2017-00171
3. Rosner W, Auchus RJ, Azziz R, Sluss PM, Raff H. Position statement: Utility, limitations, and pitfalls in measuring testosterone: an Endocrine Society position statement. J Clin Endocrinol Metab. 2007;**92**:405-413
4. Handelsman DJ, Wartofsky L. Requirement for mass spectrometry sex steroid assays in the Journal of Clinical Endocrinology and Metabolism. J Clin Endocrinol Metab. 2013;**98**:3971-3973;doi.10.1210/jc.2013-3375
5. Ly LP, Handelsman DJ. Empirical estimation of free testosterone from testosterone and sex hormone-binding globulin immunoassays. Eur J Endocrinol. 2005;**152**:471-478
6. Sartorius G, Ly LP, Sikaris K, McLachlan R, Handelsman DJ. Predictive accuracy and sources of variability in calculated free testosterone estimates. Ann Clin Biochem. 2009;**46**:137-143
7. Ly LP, Sartorius G, Hull L, Leung A, Swerdloff RS, Wang C*, et al.* Accuracy of calculated free testosterone formulae in men. Clin Endocrinol (Oxf). 2010;**73**:382-388;doi.10.1111/j.1365-2265.2010.03804.x
8. Hackbarth JS, Hoyne JB, Grebe SK, Singh RJ. Accuracy of calculated free testosterone differs between equations and depends on gender and SHBG concentration. Steroids. 2011;**76**:48-55;doi.S0039-128X(10)00229-1
9. Salameh WA, Redor-Goldman MM, Clarke NJ, Reitz RE, Caulfield MP. Validation of a total testosterone assay using high-turbulence liquid chromatography tandem mass spectrometry: total and free testosterone reference ranges. Steroids. 2010;**75**:169-175;doi.10.1016/j.steroids.2009.11.004

10. Zakharov MN, Bhasin S, Travison TG, Xue R, Ulloor J, Vasan RS*, et al.* A multi-step, dynamic allosteric model of testosterone's binding to sex hormone binding globulin. Mol Cell Endocrinol. 2015;**399**:190-200;doi.10.1016/j.mce.2014.09.001
11. Vermeulen A, Verdonck L, Kaufman JM. A critical evaluation of simple methods for the estimation of free testosterone in serum. J Clin Endocrinol Metab. 1999;**84**:3666-3672;doi.10.1210/jcem.84.10.6079
12. Hsu B, Hirani V, Naganathan V, Blyth FM, Le Couteur DG, Seibel MJ*, et al.* Sexual Function and Mortality in Older Men: The Concord Health and Ageing in Men Project. J Gerontol A Biol Sci Med Sci. 2017;**72**:520-527;doi.10.1093/gerona/glw101
13. Hsu B, Cumming RG, Hirani V, Blyth FM, Naganathan V, Le Couteur DG*, et al.* Temporal Trend in Androgen Status and Androgen-Sensitive Outcomes in Older Men. J Clin Endocrinol Metab. 2016;**101**:1836-1846;doi.10.1210/jc.2015-3810
14. Hsu B, Cumming RG, Blyth FM, Naganathan V, Le Couteur DG, Seibel MJ*, et al.* Longitudinal and cross-sectional relationships of circulating reproductive hormone levels to self-rated health and health-related quality of life in community-dwelling older men. J Clin Endocrinol Metab. 2014;**99**:1638-1647;doi.10.1210/jc.2013-3984
15. Hsu B, Cumming RG, Naganathan V, Blyth FM, Le Couteur DG, Seibel MJ*, et al.* Longitudinal relationships of circulating reproductive hormone with functional disability, muscle mass, and strength in community-dwelling older men: the Concord Health and Ageing in Men project. J Clin Endocrinol Metab. 2014;**99**:3310-3318;doi.10.1210/jc.2014-1124
16. Hsu B, Cumming RG, Naganathan V, Blyth FM, Le Couteur DG, Seibel MJ*, et al.* Associations between circulating reproductive hormones and SHBG and prevalent and incident metabolic syndrome in community-dwelling older men: the Concord Health and Ageing in Men Project. J Clin Endocrinol Metab. 2014;**99**:E2686-2691;doi.10.1210/jc.2014-2464
17. Hsu B, Cumming RG, Blyth FM, Naganathan V, Le Couteur DG, Seibel MJ*, et al.* The longitudinal relationship of sexual function and androgen status in older men: the Concord Health and Ageing in Men Project. J Clin Endocrinol Metab. 2015;**100**:1350-1358;doi.10.1210/jc.2014-4104
18. Hsu B, Cumming RG, Seibel MJ, Naganathan V, Blyth FM, Bleicher K*, et al.* Reproductive Hormones and Longitudinal Change in Bone Mineral Density and Incident Fracture Risk in Older Men: The Concord Health and Aging in Men Project. J Bone Miner Res. 2015;**30**:1701-1708;doi.10.1002/jbmr.2493
19. Hsu B, Cumming RG, Waite LM, Blyth FM, Naganathan V, Le Couteur DG*, et al.* Longitudinal Relationships between Reproductive Hormones and Cognitive Decline in Older Men: The Concord Health and Ageing in Men Project. J Clin Endocrinol Metab. 2015;**100**:2223-2230;doi.10.1210/jc.2015-1016
20. Yeap BB, Alfonso H, Chubb SA, Hankey GJ, Handelsman DJ, Golledge J*, et al.* In older men, higher plasma testosterone or dihydrotestosterone is an independent predictor for reduced incidence of stroke but not myocardial infarction. J Clin Endocrinol Metab. 2014;**99**:4565-4573;doi.10.1210/jc.2014-2664
21. Yeap BB, Alfonso H, Chubb SA, Gauci R, Byrnes E, Beilby JP*, et al.* Higher serum undercarboxylated osteocalcin and other bone turnover markers are associated with

497 reduced diabetes risk and lower estradiol concentrations in older men. J Clin Endocrinol
498 Metab. 2015;**100**:63-71;doi.10.1210/jc.2014-3019
499 22. Yeap BB, Knuiman MW, Divitini ML, Hui J, Arscott GM, Handelsman DJ*, et al.*
500 Epidemiological and Mendelian Randomization Studies of Dihydrotestosterone and
501 Estradiol and Leukocyte Telomere Length in Men. J Clin Endocrinol Metab.
502 2016;**101**:1299-1306;doi.10.1210/jc.2015-4139
503 23. Yeap BB, Alfonso H, Chubb SA, Handelsman DJ, Hankey GJ, Almeida OP*, et al.* In
504 older men an optimal plasma testosterone is associated with reduced all-cause
505 mortality and higher dihydrotestosterone with reduced ischemic heart disease
506 mortality, while estradiol levels do not predict mortality. J Clin Endocrinol Metab.
507 2014;**99**:E9-18;doi.10.1210/jc.2013-3272
508 24. Yeap BB, Alfonso H, Paul Chubb SA, Hankey GJ, Handelsman DJ, Golledge J*, et al.*
509 In older men, higher plasma testosterone or dihydrotestosterone are independent
510 predictors for reduced incidence of stroke but not myocardial infarction. J Clin
511 Endocrinol Metab. 2014:jc20142664;doi.10.1210/jc.2014-2664
512 25. Yeap BB, Alfonso H, Chubb SA, Handelsman DJ, Hankey GJ, Norman PE*, et al.*
513 Reference Ranges and Determinants of Testosterone, Dihydrotestosterone, and
514 Estradiol Levels Measured using Liquid Chromatography-Tandem Mass Spectrometry in
515 a Population-Based Cohort of Older Men. J Clin Endocrinol Metab. 2012;**97**:4030-
516 4039;doi.10.1210/jc.2012-2265
517 26. Cumming RG, Handelsman D, Seibel MJ, Creasey H, Sambrook P, Waite L*, et al.*
518 Cohort Profile: the Concord Health and Ageing in Men Project (CHAMP). Int J Epidemiol.
519 2009;**38**:374-378;doi.dyn071
520 27. Harwood DT, Handelsman DJ. Development and validation of a sensitive liquid
521 chromatography-tandem mass spectrometry assay to simultaneously measure
522 androgens and estrogens in serum without derivatization. Clin Chim Acta. 2009;**409**:78-
523 84
524 28. Keski-Rahkonen P, Desai R, Jimenez M, Harwood DT, Handelsman DJ.
525 Measurement of Estradiol in Human Serum by LC-MS/MS Using a Novel Estrogen-
526 Specific Derivatization Reagent. Anal Chem. 2015;**87**:7180-
527 7186;doi.10.1021/acs.analchem.5b01042
528 29. Ware JE, Kosinski DM, Turner-Bowker DM, Sundaram M, Gandek B, Maruish ME.
529 User's Manual for the SF-12v2 Health Survey. 2nd ed.; 2009.
530 30. Smith LA, Branch LG, Scherr PA, Wetle T, Evans DA, Hebert L*, et al.* Short-term
531 variability of measures of physical function in older people. J Am Geriatr Soc.
532 1990;**38**:993-998
533 31. Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J*, et al.* Frailty
534 in older adults: evidence for a phenotype. J Gerontol A Biol Sci Med Sci. 2001;**56**:M146-
535 156
536 32. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for
537 grading the cognitive state of patients for the clinician. J Psychiatr Res. 1975;**12**:189-198
538 33. Shiekh J, Yesavage JA. Geriatric depression scale: recent findings and
539 developments of a short version. QualityMetric Inc; 2009.

34. O'Donnell AB, Araujo AB, Goldstein I, McKinlay JB. The validity of a single-question self-report of erectile dysfunction. Results from the Massachusetts Male Aging Study. J Gen Intern Med. 2005;**20**:515-519
35. Grundy SM, Brewer HB, Jr., Cleeman JI, Smith SC, Jr., Lenfant C. Definition of metabolic syndrome: Report of the National Heart, Lung, and Blood Institute/American Heart Association conference on scientific issues related to definition. Circulation. 2004;**109**:433-438;doi.10.1161/01.cir.0000111245.75752.c6
36. Washburn RA, McAuley E, Katula J, Mihalko SL, Boileau RA. The physical activity scale for the elderly (PASE): evidence for validity. J Clin Epidemiol. 1999;**52**:643-651
37. Heymsfield SB, Smith R, Aulet M, Bensen B, Lichtman S, Wang J*, et al.* Appendicular skeletal muscle mass: measurement by dual-photon absorptiometry. Am J Clin Nutr. 1990;**52**:214-218
38. Studenski SA, Peters KW, Alley DE, Cawthon PM, McLean RR, Harris TB*, et al.* The FNIH sarcopenia project: rationale, study description, conference recommendations, and final estimates. J Gerontol A Biol Sci Med Sci. 2014;**69**:547-558;doi.10.1093/gerona/glu010
39. Hsu B, Cumming RG, Naganathan V, Blyth FM, Le Couteur DG, Hirani V*, et al.* Temporal Changes in Androgens and Estrogens Are Associated With All-Cause and Cause-Specific Mortality in Older Men. J Clin Endocrinol Metab. 2016;**101**:2201-2210;doi.10.1210/jc.2016-1025
40. Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. Biometrics. 1986;**42**:121-130
41. Kaufman JM, Vermeulen A. The decline of androgen levels in elderly men and its clinical and therapeutic implications. Endocr Rev. 2005;**26**:833-876
42. Bhasin S, Cunningham GR, Hayes FJ, Matsumoto AM, Snyder PJ, Swerdloff RS*, et al.* Testosterone therapy in men with androgen deficiency syndromes: an Endocrine Society clinical practice guideline. J Clin Endocrinol Metab. 2010;**95**:2536-2559;doi.95/6/2536
43. Sodergard R, Backstrom T, Shanbhag V, Carstensen H. Calculation of free and bound fractions of testosterone and estradiol-17 beta to human plasma proteins at body temperature. J Steroid Biochem. 1982;**16**:801-810
44. Mazer NA. A novel spreadsheet method for calculating the free serum concentrations of testosterone, dihydrotestosterone, estradiol, estrone and cortisol: with illustrative examples from male and female populations. Steroids. 2009;**74**:512-519
45. Antonio L, Wu FC, O'Neill TW, Pye SR, Ahern TB, Laurent MR*, et al.* Low Free Testosterone Is Associated with Hypogonadal Signs and Symptoms in Men with Normal Total Testosterone. J Clin Endocrinol Metab. 2016;**101**:2647-2657;doi.10.1210/jc.2015-4106
46. Wu FC, Tajar A, Beynon JM, Pye SR, Silman AJ, Finn JD*, et al.* Identification of late-onset hypogonadism in middle-aged and elderly men. N Engl J Med. 2010;**363**:123-135;doi.NEJMoa0911101
47. Kelleher S, Conway AJ, Handelsman DJ. Blood testosterone threshold for androgen deficiency symptoms. J Clin Endocrinol Metab. 2004;**89**:3813-3817

583 48. Zitzmann M, Faber S, Nieschlag E. Association of specific symptoms and
584 metabolic risks with serum testosterone in older men. J Clin Endocrinol Metab.
585 2006;**91**:4335-4343
586 49. Holden CA, McLachlan RI, Cumming R, Wittert G, Handelsman DJ, de Kretser DM*,*
587 *et al.* Sexual activity, fertility and contraceptive use in middle-aged and older men: Men
588 in Australia, Telephone Survey (MATeS). Hum Reprod. 2005;**20**:3429-3434;doi.dei307
589 50. Bremner WJ, Vitiello MV, Prinz PN. Loss of circadian rhythmicity in blood
590 testosterone levels with aging in normal men. J Clin Endocrinol Metab. 1983;**56**:1278-
591 1281

**TABLE 1. Characteristics of the study health outcomes at baseline, 2-year and 5-year**

| | Baseline (n=1651) Mean (SD) or N (%) | 2-year (n=1291) Mean (SD) or N (%) | 5-year (n=910) Mean (SD) or N (%) | Non-participation at 2-year‡ (n=345) | Non-participation at 5-year‡ (n=747) |
|---|---|---|---|---|---|
| Age (years) | 76.9 ± 5.5 | 79.0 ± 25.8 | 81.4 ± 4.6 | 79.1 ± 6.1 | 78.8 ± 6.0 |
| Comorbidity | 2.6 ± 1.8 | 2.5 ± 1.7 | 2.5 ± 1.6 | 2.9 ± 1.9 | 2.9 ± 1.9 |
| BMI ($kg/m^2$) | 27.8 ± 1.8 | 27.8 ± 4.0 | 27.6 ± 4.0 | 27.5 ± 4.2 | 27.6 ± 4.3 |
| MMSE | 27.1 ± 3.05 | 27.4 ± 2.8 | 27.2 ± 3.1 | 26.1 ± 3.7 | 26.4 ± 3.4 |
| PASE | 124.4 ± 62.1 | 119.8 ± 59.7 | 117.4 ± 63.2 | 100.6 ± 62.8 | 107.9 ± 61.2 |
| Walking speed (m/s) | 0.9 ± 0.2 | 0.9 ± 0.2 | 0.9 ± 0.2 | 0.8 ± 0.2 | 0.8 ± 0.2 |
| Hip BMD ($g/cm^2$) | 0.9 ± 0.1 | 0.9 ± 0.1 | 0.9 ± 0.1 | 0.9 ± 0.2 | 0.9 ± 0.2 |
| SF-12 Physical | 48.6 ± 10.5 | 48.6 ± 10.5 | 47.6 ± 10.6 | 45.7 ± 11.7 | 46.1 ± 11.1 |
| SF-12 Mental | 49.1 ± 6.4 | 49.3 ± 6.3 | 49.6 ± 6.5 | 48.3 ± 7.0 | 48.6 ± 7.0 |
| Weight (kg) | 79.4 ± 13.0 | 79.2 ± 12.8 | 78.1 ± 12.7 | 77.5 ± 13.5 | 77.9 ± 13.5 |
| Grip strength (kg) | 34.5 ± 7.5 | 34.7 ± 8.0 | 32.7 ± 8.3 | 32.0 ± 7.6 | 32.5 ± 7.0 |
| Lean mass (kg) | 7.2 ± 1.2 | 7.2 ± 1.2 | 7.2 ± 1.2 | 7.0 ± 1.3 | 7.1 ± 1.2 |
| Fat percentage (%) | 28.9 ± 6.0 | 29.2 ± 6.0 | 29.7 ± 6.1 | 28.9 ± 6.4 | 29.0 ± 6.2 |
| Glucose (mmol/L) | 5.6 ± 1.4 | 5.7 ± 1.5 | 5.7 ± 1.5 | 5.6 ± 1.4 | 5.6 ± 1.5 |
| Hemoglobin (g/dL) | 142.9 ± 14.0 | 142.2 ± 13.5 | 141.0 ± 14.4 | 138.6 ± 17.0 | 140.2 ± 15.8 |
| Waist (cm) | 103.4 ± 19.9 | 102.1 ± 11.2 | 101.1 ± 11.0 | 104.3 ± 39.1 | 103.6 ± 27.7 |
| Current Smoker | 101 (6%) | 52 (4%) | 35 (4%) | 32 (9%) | 49 (7%) |
| Poor Self-rated Health | 500 (30%) | 389 (29%) | 251 (26%) | 132 (41%) | 272 (38%) |
| ADL disability | 138 (8%) | 141 (10%) | 119 (13%) | 69 (20%) | 104 (14%) |
| Frail | 158 (10%) | 129 (10%) | 93 (10%) | 74 (23%) | 130 (18%) |
| Previous Falls | 138 (8%) | 125 (15%) | 115 (12%) | 47 (14%) | 84 (11%) |
| Depression | 242 (15%) | 206 (15%) | 122 (13%) | 86 (26%) | 159 (22%) |
| Erectile dysfunction | 441 (36%) | 306 (33%) | 121 (20%) | 62 (31%) | 153 (32%) |
| No sexual activity | 532 (44%) | 379 (41%) | 241 (41%) | 65 (33%) | 143 (31%) |
| Low sexual satisfaction | 1031 (89%) | 783 (90%) | 509 (90%) | 161 (85%) | 368 (86%) |
| Low sexual desire | 371 (30%) | 276 (30%) | 179 (30%) | 65 (33%) | 133 (28%) |
| Metabolic syndrome | 481 (37%) | 472 (40%) | 474 (53%) | 90 (36%) | 202 (36%) |

* BMI: body mass index, MMSE: mini mental status examination, PASE: physical activity scale for the elderly, BMD: bone mineral density, ADL: activities daily of living

†Higher values are better for MMSE (out of 30), PASE, SF-12 Physical and Mental (each out of 100). Lower values are better for comorbidity (out of 19).

‡The data for non-participation at 2-year and 5-year are their baseline descriptive characteristic. Death was the main reason for non-participation at 2 years (99 deaths) and at 5 years (382 deaths).

**TABLE 2. Serum testosterone (T), SHBG and free testosterone (cFT) levels for the CHAMP cohort at baseline according to different T/cFT status cutoff and calculation**

| | N (%) | T ((nmol/L) Mean (SD) | SHBG (nmol/L) Mean (SD) | cFT (pmol/L) Mean (SD) |
|---|---|---|---|---|
| CHAMP cutoff and calculation* | | | | |
| All | 1651, 100% | 14.7 ± 6.4 | 50.1 ± 20.7 | 206.6 ± 78.0 |
| Normal T/Normal cFT (NN) | 1283, 78% | 17.0 ± 5.1 | 52.3 ± 20.2 | 235.5 ± 56.8 |
| Normal T/Low cFT (NL) | 40, 2% | 11.2 ± 0.7 | 61.9 ± 17.2 | 149.2 ± 5.0 |
| Low T/Normal cFT (LN) | 38, 2% | 9.4 ± 0.8 | 24.3 ± 5.9 | 164.1 ± 8.2 |
| Low T/Low cFT (LL) | 290, 18% | 5.9 ± 3.4 | 42.4 ± 20.6 | 91.8 ± 52.6 |

*T level below or above 10.2 nmol/L and cFT level below or above 156 pmol/L (lowest quintile)

**TABLE 3. Summary of the concordant, discordant and isolated discordance for the morbidity outcomes**

| | Primary* | | Sensitivity 1* | | Sensitivity 2* | |
|---|---|---|---|---|---|---|
| **Cross-sectional** | **Unadjusted** | **Adjusted** | **Unadjusted** | **Adjusted** | **Unadjusted** | **Adjusted** |
| Concordant (only significant LL)† | 19 | 15 | 17 | 12 | 17 | 12 |
| Discordant (both significant LN & significant LL) | 8 | 4 | 0 | 3 | 0 | 3 |
| Discordant (both significant NL & significant LL) | 8 | 3 | 7 | 3 | 7 | 3 |
| Isolated Discordance (significant LN or NL without significant LL) | 1 | 1 | 1 | 1 | 1 | 1 |
| **Longitudinal** | **Unadjusted** | **Adjusted** | **Unadjusted** | **Adjusted** | **Unadjusted** | **Adjusted** |
| Concordant (only significant LL) | 18 | 16 | 17 | 15 | 17 | 15 |
| Discordant (both significant LN & significant LL) | 7 | 6 | 13 | 8 | 13 | 8 |
| Discordant (both significant NL & significant LL) | 10 | 4 | 10 | 7 | 10 | 7 |
| Isolated Discordance (significant LN or NL without significant LL) | 2 | 0 | 1 | 1 | 1 | 1 |

*Primary; primary analysis using an empirical formula (FTZ) categorizing low T and low cFT based on lowest quintile. Sensitivity 1; sensitivity analysis using the same FTZ formula categorizing low T and low cFT based on lowest centile. Sensitivity 2; sensitivity analysis using the Vermeulen formula categorizing low T and low cFT based on lowest quintile.

†LL is Low T/ Low cFT; LN is Low T/ Normal cFT; NL is Normal T/ Low cFT; reference group is NN Normal T/ Normal cFT

**FIGURE 1. Summary of the cross-sectional and longitudinal findings for the morbidity outcomes**