TESTOSTERONE TREATMENT AND COGNITIVE FUNCTION IN OLDER MEN WITH LOW TESTOSTERONE AND AGE-ASSOCIATED MEMORY IMPAIRMENT

On-line Only Supplement 2 Materials

Contents:

ROLES OF INVESTIGATORS AND OTHERS IN THE TESTOSTERONE TRIALS

Steering Committee	
Peter J. Snyder, MD, Chair	University of Pennsylvania
Elizabeth Barrett-Connor, MD	University of California, San Diego
Shalendar Bhasin, MD	Harvard Medical School
Jane A Cauley, DrPH	University of Pittsburgh
David Cella, PhD	Northwestern University
Jill P Crandall, MD	Albert Einstein College of Medicine
Glenn Cunningham, MD	Baylor Medical School
Susan S Ellenberg, PhD	University of Pennsylvania
Kristine E Ensrud, MD, MPH	University of Minnesota
John T Farrar, MD, PhD	University of Pennsylvania
Thomas M Gill, MD	Yale University
Cora E Lewis, MD, MSPH	University of Alabama at Birmingham
Alvin M Matsumoto, MD	University of Washington
Mark E Molitch, MD	Northwestern University
Marco Pahor, MD	University of Florida
Susan Resnick, PhD	National Institute of Aging
Raymond C Rosen, PhD	New England Research Institute
Ronald S Swerdloff, MD	University of California, Los Angeles

Statisticians (Department of Biostatistics and Epidemiology, University of Pennsylvania)	
Susan S Ellenberg, PhD	
Xiaoling Hou, MS	Alisa Stephens-Shields, PhD
J Richard Landis, PhD	Bret Zeldow, MS
Liyi Cen, MD	Renee H Moore, PhD

Data Coordinating Center (Clinical Research Computing Unit, Department of Biostatistics and Epidemiology, University of Pennsylvania)	
Denise Cifelli, MS	Darlene Dougar, MPH
Laura Fluharty, MPH	Laura Gallagher, MPH, CCRP
Shawn Ballard, MS	Tracy Chai, MS
James Dattilo, BS	Trina Brown
Sandra Smith, AS	Fran Chicchi, BS

Adjudicators of Cardiovascular and Cerebrovascular Adverse Events	
Scott E Kasner, MD	University of Pennsylvania
Cora E Lewis, MD, MSPH	University of Alabama at Birmingham
Steven R Messe, MD	University of Pennsylvania
Emile R Mohler III, MD	University of Pennsylvania

Clinical Trial Sites

Site	PI	Co-Is and Sub-Is	Research Coordinator(s)
Albert Einstein College of Medicine	Jill P. Crandall, MD	Vafa Tabatabaie, MD Eric Epstein, MD Uriel Barzel, MD	Gilda Trandafirescu
Baylor College of Medicine	Glenn R. Cunningham, MD	N/A	Emilia Cordero, MS, RN, ANP-C Patti Marino
Brigham and Women's Hospital	Shalender Bhasin, MD	Shehzad Basaria, MD	Richard Eder Erica Appleman Kathleen Ann Halley
Northwestern University	Mark Molitch, MD	Daniel Toft, MD Amisha Wallia, MD	Diane Larsen Elaine Massaro, MS, RN Daphne Adelman, BSN, MBA
University of Alabama at Birmingham	Cora E. Lewis, MD, MSPH	James Shikany, DrPH Peter Kolettis, MD	Phillip Johnson Margaret N. Pike, RN Isabelle J. Joffrion, RN
University of California at Los Angeles	Ronald Swerdloff, MD	Christina Wang, MD	Xiaodan Han, RN Jamila Ashai
University of California at San Diego	Elizabeth Barrett-Connor, MD	Karen Herbst, MD Heather Hofflich, DO J. Kellogg Parsons, MD	Noralinda Kamantigue, BSN, RN Mary Lou Carrion-Peterson, BSN, RN Gabriela Reno Lauren Claravall Jean Smith, RN
University of Florida	Marco Pahor, MD	Susan Nayfield, MD, M.Sc Stephen D. Anton, Ph.D. Todd Manini, Ph.D Philip Dahm, MD Michael Marsiske, Ph.D Bhanuprasad Sandesara, MD	Melissa Lewis Mieniecia L. Black, MPH Jeffrey Knaggs William Marena Jane Ching-ju Lu, MPH
University of Minnesota	Kristine E. Ensrud, MD, MPH	Susan J. Diem, MD, MPH Howard Fink, MD, MPH Christopher Warlick, MD	Sandra Potter, CCRC Luanne Welch, RN Pamela Van Covering, MA Kristi Lee Jacobson Lisa Miller
University of Pittsburgh	Jane Cauley, DrPH	Mara J. Horwitz, MD Susan L. Greenspan, MD Thomas M. Jaffe, MD	Linda Prebehalla, RN Janet T. Bonk, RN, MPH Jennifer L. Rush, MPH
University of Washington	Alvin M. Matsumoto, MD	N/A	Janet Gilchriest, RN Kathy Winter Magdalena Wojtowicz, RN
Yale University	Thomas M. Gill, MD	Natalie deRekeneire, MD Susan Kashaf, MD Lee Katz, MD Hamid Mojibian, MD	Joanne McGloin Karen Wu Dismayra Martinez, MA Denise Shepard, BSN, MBA

Data and Safety Monitoring Board

Member	Affiliation
Michael O. Thorner, MD	Department of Medicine, University of Virginia
Eric Klein, MD	Cleveland Clinic
Wayne A. Meikle, MD	University of Utah
Wayne J. Hellstrom, MD	Department of Urology, Tulane University
J. Philip Miller, PhD (Chair)	Washington University
Richard Chappell, PhD	Department of Biostatistics, University of Wisconsin
George A. Kuchel, MD	Division of Geriatrics, University of Connecticut
Manuel D. Cerqueira, MD	Cleveland Clinic Nuclear Medicine
E. Magnus Ohman, MD	Duke University Medical Center

TITLE: THE TESTOSTERONE TRIAL

PROTOCOL

Sponsors	**National Institute on Aging (NIA) and AbbVie, Inc.**
NIH Grant Number	**U01 AG030644**
University of Pennsylvania Protocol Number	**808676**
Principal Investigator	**Peter J. Snyder, MD**
Study Drug Provider	**AbbVie Inc**
IND Number	**104707**
Clinical Trials.gov Number	**NCT00799617**

Date:	**October 6, 2008**	**Version Number: 1.0**
Amended:	**December 22, 2008**	**Version Number: 1.2**
Amended:	**April 1, 2009**	**Version Number: 1.3**
Amended:	**June 4, 2009**	**Version Number 1.4**
Administrative Changes:	**September 15, 2009**	**Version Number 1.5**
Amended:	**April 8, 2010**	**Version Number 2.0**
Amended:	**June 22, 2010**	**Version Number 3.0**
Amended:	**September 13, 2010**	**Version Number 3.1**
Amended:	**November 23, 2010**	**Version Number 3.2**
Amended:	**April 19, 2011**	**Version Number 3.3**
Amended:	**June 23, 2011**	**Version Number 3.4**
Amended:	**August 1, 2011**	**Version Number 3.5**
Amended:	**December 7, 2011**	**Version Number 3.6**
Amended:	**February 7, 2012**	**Version Number 3.7**
Amended:	**March 1, 2012**	**Version Number 3.8**
Amended:	**April 16, 2012**	**Version Number 4.0**
Amended:	**August 13, 2012**	**Version Number 5.0**

Amended:	**October 17, 2012**	**Version 6.0**
Amended:	**October 24, 2012**	**Version 7.0**
Amended:	**January 23, 2013**	**Version 7.1**
Amended:	**March 15, 2013**	**Version 7.2**
Amended:	**July 16, 2013**	**Version 7.3**

TABLE OF CONTENTS

Study Summary

Title	The Testosterone Trials
Protocol Number	808676
Study Design	Randomized, placebo-controlled, double-blind study of five coordinated trials
Study Duration	Six years
Study Centers	Multi-center set of trials involving 12 clinical sites geographically distributed across the United States
Objectives	The primary specific aims are to test the hypotheses that testosterone treatment of elderly men whose serum testosterone concentrations are unequivocally low – and who have symptoms and/or objectively measured abnormalities in at least one of five areas that could be due to low testosterone (physical or sexual function, vitality, cognition, and anemia) – will result in more favorable changes in those abnormalities than placebo treatment.
Number of Subjects	800
Diagnosis and Main Inclusion Criteria	A constellation of conditions that occur as men age will be studied: mobility disability, decreased libido, low vitality, reduced memory performance, as well as anemia, all of which could be at least partially the result of low testosterone. Primary entry criteria will be age ≥65 years, an unequivocally low testosterone concentration (average of 2 morning testosterone values, < 275 ng/dL), and symptoms and objective manifestations of mobility disability, low libido, or low vitality.
Study Product, Dose, Route, Regimen	AndroGel®, testosterone in an alcohol-water gel, will be administered transdermally in doses from 5 to 15 grams per day, as necessary to maintain the serum testosterone concentration within the range of normal for young men.
Duration of administration	AndroGel or placebo will be administered to each subject for 12 months.
Reference therapy	The effects of AndroGel on the primary and secondary end points will be compared to effects of placebo on these end points.
Statistical Methodology	The primary end points for each of the five trials (Physical and Sexual Function, Vitality, Cognitive Function and Anemia) will be analyzed separately by random effects models for each specific trial.

1. Introduction

This document is a protocol for a human research study to be conducted according to US and international standards of Good Clinical Practice (FDA Title 21 part 312 and International Conference on Harmonization guidelines), applicable government regulations and institutional research policies and procedures.

This trial is supported by the National Institute on Aging (NIA), the National Institute of Neurological Disorders and Stroke (NINDS), the National Institute of Child Health and Human Development (NICHD), the National Heart, Lung and Blood Institute (NHLBI) and AbbVie , Inc.

1.1. Background

As men get older, they experience many conditions, often together, that eventually result in the inability to perform many activities of daily living, an increased propensity to fall, and decreased independence. These conditions include mobility disability and low vitality. Elderly men also experience increased anemia, metabolic syndrome, decreased sexual function and memory impairment. These conditions likely have multiple causes, but one cause that could contribute to all of them is a low serum testosterone concentration. When young hypogonadal men are treated with testosterone, they experience improvements in sexual function, muscle mass and strength, bone mineral density, sense of well being, and anemia. However, the benefits of testosterone therapy in older men with age-related decline in testosterone concentration are not known and are the subject of this investigation.

1.2. Decrease in Testosterone as Men Age

As men age, their serum testosterone concentration falls gradually from age 20 to over age 80, as demonstrated by both cross-sectional (1) and longitudinal studies (2-4). By the eighth decade, approximately 30% of men have concentrations of total testosterone lower than normal for young men and 70% have free testosterone concentrations lower than normal for young men (3). Age-related decline in testosterone concentrations is associated with decreases in physical function, sexual function, vitality, and, in some studies, decreases in memory and cognitive function.

1.3. Conditions that Testosterone Treatment Might Improve

1.3.1. Physical Function

As men age, they experience a decrease in muscle mass and strength and in physical function (5). Decreased muscle mass and strength leads to impairment of physical function and mobility (6, 7). Mobility disability is a highly prevalent recognized geriatric syndrome. The 6-minute walk test, which assesses walking speed and distance, is a standardized, reliable measure of mobility (8).

In population studies in elderly men (9, 10), lower testosterone concentrations are associated with decreased physical function. Testosterone treatment of young hypogonadal men significantly increases lean body mass (11, 12) and muscle strength (11). Clinical trials in which elderly men with low-normal serum testosterone concentrations were treated with testosterone have consistently demonstrated an increase in muscle mass, but less consistently demonstrated increases in muscle strength and physical function (13-15). Limited data from clinical trials suggest that testosterone therapy might improve walking speed.

1.3.2. Sexual Function

Aging in men is associated with reduced sexual activity, which may respond to testosterone. Sexual desire, erection, and ejaculation decrease linearly from 20 to 70 years (16-18). Erectile dysfunction occurs in approximately 20-30% of men in their 50s, and by age 70, most men have lost the capacity for firm erection or satisfying orgasm. One possible explanation for the decline in sexual function with age is the concomitant fall in testosterone. A meta-analysis of randomized, placebo-controlled studies concluded that testosterone improves sexual function, the more so the lower the pretreatment testosterone concentrations (19). The possibility that testosterone treatment will improve sexual function may depend on other factors, such as the availability of a willing partner, use of PDE5 inhibitors, and overall health. It is also possible that testosterone will affect some aspects of sexual function more than others, especially Hypoactive Sexual Desire Disorder (diminished libido).

1.3.3. Vitality

Several lines of evidence suggest that a decrease in testosterone contributes to age-related decreases in vitality, sense of well-being, and quality of life. Several epidemiologic studies have documented an association between serum testosterone concentrations and mood and vitality, mostly in the setting of depression (20, 21). Low testosterone in elderly men is associated more with subsyndromal depression and related symptoms than with major depression. The ability of testosterone treatment to improve vitality, mood, and well being in men who are severely hypogonadal due to known pituitary or testicular disease is accepted by endocrinologists. A few prospective studies have also documented improvements in vitality and well being during testosterone therapy (22-24).

1.3.4. Cognitive Function

The fall in testosterone levels with increasing age is accompanied by a decline in cognitive function, including reductions in verbal and visual memory, spatial ability and executive function (25-27). Several studies of elderly men suggest that age-related declines in circulating testosterone levels are associated with reduced cognitive function, and elderly men with prostate cancer made hypogonadal by androgen deprivation therapy show cognitive impairments relative to their pretreatment performance in verbal memory, visual memory, spatial ability, and executive function (28-30). In addition, a number of small randomized trials suggest that testosterone may benefit memory in elderly men. Together, these studies suggest that lower testosterone levels or androgen action are associated with poorer cognitive functioning in otherwise healthy elderly men and that testosterone treatment may improve memory functioning.

1.3.5. Anemia

Testosterone is well known to stimulate erythropoiesis (12). Low testosterone is associated with anemia in elderly men (31). Anemia in elderly men is associated with current disability (32) and predicts future morbidity and mortality after adjusting for comorbidities (33). We shall therefore determine if testosterone corrects anemia.

1.4. Conditions Testosterone Might Worsen

1.4.1. Prostate Cancer

Elderly men often harbor clinically silent prostate cancer. The testosterone-dependency of metastatic prostate cancer is illustrated by regression following surgical or medical castration

(34) and exacerbation following testosterone treatment (35). Many elderly men harbor occult prostate cancer (36). There is no direct evidence, however, that either high endogenous serum testosterone concentrations or testosterone treatment of men with low testosterone concentrations increases the risk of clinical prostate cancer. Randomized, placebo-controlled trials of testosterone in elderly men show that testosterone increases PSA but not prostate cancer, although their statistical power to detect a difference between treatment groups was very small (37).

One challenge with regard to prostate cancer in planning this trial is to protect individuals who volunteer. We shall exclude men who have a prostate nodule by manual examination or a serum PSA concentration above a defined value; and then we will monitor men who do enroll by repeating the manual examination and the PSA measurement during the trial. Unfortunately, there is no PSA value that has both high specificity and sensitivity for detecting prostate cancer (38). A Prostate Cancer Risk Calculator was devised to allow prediction of a man's risk of both overall and high-grade prostate cancer. This Risk Calculator (http://www.compass.fhcrc.org/edrnnci/bin/calculator/main.asp) has been applied to another population of 3488 men (39). Because this Calculator does not take into account the serum testosterone concentration, and a low serum testosterone concentration results in a lower PSA, we shall adjust the serum PSA concentration to account for the low testosterone. This adjustment will be based on the regression coefficient (0.00128) derived from data from the European Male Aging Study showing a direct correlation between serum testosterone and PSA. Each man's PSA will be adjusted to what it would be if his serum testosterone were 460 ng/dL as in the following equation: Adjusted PSA = PSA + (460 - testosterone level) x 0.00128. For example, if a man's measured PSA is 1.0 ng/mL and testosterone is 200 ng/dL, the PSA will be adjusted upward by (460 - 200) x 0.00128 = 0.33. The adjusted PSA, 1.33 ng/mL, would then be used in the Prostate Cancer Risk Calculator. The serum PSA will also be increased, and specifically, doubled, before its use, when the subject is taking a 5-alpha reductase inhibitor.

In addition to selecting men at relatively low risk of developing prostate cancer by using the Prostate Risk Calculator, we have proposed criteria by which to monitor them during a testosterone trial. We chose a PSA increment criterion based on data from the placebo arm of a finasteride study. Taking into account the upward adjustment of the baseline PSA of 0.3-0.4 ng/mL, as above, we shall use an increment of 1.0 ng/mL above the adjusted baseline PSA confirmed by repeat determination, as the criterion for referral for urological evaluation and prostate biopsy.

1.4.2. Benign Prostatic Hypertrophy (BPH) and Lower Urinary Tract Symptoms (LUTS)

Despite the theoretical reasons that testosterone treatment could increase the risk of LUTS due to BPH, interventional studies have not demonstrated this risk. We shall monitor lower urinary tract symptoms during this trial.

1.4.3. Erythrocytosis

Testosterone stimulates erythropoiesis, so a potential consequence is erythrocytosis. We shall determine if a man whose hemoglobin is normal before treatment experiences an increase above normal (erythrocytosis) during treatment.

1.4.4. Sleep Apnea

Some evidence suggests that testosterone may exacerbate sleep apnea, although the evidence is weak. To be safe, we shall exclude men with diagnosed but untreated sleep apnea from these trials.

1.4.5. Cardiovascular Disease

In a recent study in men ≥65 years of age, men treated with testosterone experienced significantly more cardiac serious adverse events than men treated with placebo (unpublished). However, in another recent study (Srinivas-Shankar, J Clin Endocrinol Metab 95: 1220, 2010), no such excess occurred. In addition, Murad and Montori performed a meta-analysis of 51 randomized controlled trials of testosterone, in which nine reported serious cardiac adverse events. The quality of evidence was considered low because of few events, brief length of observation, and substantial loss of subjects to observation. Nonetheless, they concluded that "Compared with placebo, testosterone therapy was not associated with a significant increase in the risk of death, myocardial infarction (MI), revascularization procedures, cardiac arrhythmias, or a cardiac composite that included MI, revascularization procedures and cardiac arrhythmias." Including these new trials did not change the conclusions.

1.5. Genetic Propensity to Respond to Testosterone

Variability in both beneficial and deleterious effects of testosterone may be explained by the fact that serum testosterone concentrations do not predict androgen responsiveness well, most likely due to genetic differences in 1) androgen action or metabolism due to relative conversion to active metabolites (such as estradiol or dihydrotestosterone), binding proteins (i.e. SHBG), and/or tissue-specific coactivators or corepressors; or 2) tissue-specific end-organ response to androgens due to coexistent polymorphisms of modulator genes. Our strategy is to evaluate this genetic predilection to beneficial and adverse effects of treatment by testosterone. Therefore, we shall collect peripheral blood lymphocytes for genetic analyses.

2. Study Objectives

The primary specific aims of the coordinated set of randomized, placebo-controlled clinical trials are to test the hypotheses that testosterone treatment of elderly men whose serum testosterone concentrations are unequivocally low – and who have symptoms and/or objectively measured abnormalities that could be due to low testosterone (physical or sexual function, vitality, cognition, or anemia) – will result in more favorable changes in those abnormalities than placebo treatment. The trials are highly coordinated, but each trial has its own primary, secondary, and exploratory specific aims, as follows:

2.1. Physical Function Trial

Primary specific aim: To test the hypothesis that testosterone treatment for one year, compared with placebo, of men ≥65 years who have an average serum testosterone concentration < 275 ng/dL and mobility disability, as defined by self-reported difficulty in walking 1/4 mile and objectively measured gait speed <1.2 meters/second on the six-minute walk test, will be associated with a greater proportion of men improving their six-minute walking distance by >50 m.

Secondary specific aim: To test the hypotheses that testosterone treatment of these same men for one year, compared with placebo treatment, will be associated with greater improvement in self-reported physical function by the 10-item physical function (PF10) component of the SF36.

Exploratory aims: To determine if testosterone treatment, compared with placebo, will be associated with

1. Better patient global impression of change in walking ability,

2. A greater proportion of men in all of the trials (combined) improving their six-minute walking distance >50m,
3. A lower frequency of falls in men in this Trial and in all trials.

2.2. Sexual Function Trial

Primary specific aim: To test the hypothesis that testosterone treatment for one year, compared with placebo, of men ≥65 years who have an average serum testosterone concentration < 275 ng/dL and decreased libido by self-report and by the Derogatis Interview for Sexual Functioning in Men-II (DISF-M-II) questionnaire, will be associated with greater improvement in sexual activity, as assessed by the Harbor-UCLA 7-day Sexual Function Questionnaire, question 4.

Secondary specific aims: To test the hypotheses that in these men, testosterone treatment for one year, compared with placebo treatment, will be associated with more favorable outcomes in

1. Harbor-UCLA 7-day Sexual Function Questionnaire, Questions 1–3, and 5-6,
2. Libido, as assessed by the DISF-M-II,
3. Erectile function as assessed by International Index of Erectile Function (IIEF).

Exploratory aims: To determine if testosterone treatment for one year, compared with placebo, will be associated with

1. Better patient global impression of change in sexual activity,
2. More favorable change in the UCLA 7-day Sexual Function Questionnaire among men in all trials combined.

2.3. Vitality Trial

Primary specific aim: To test the hypothesis that testosterone treatment for one year, compared with placebo treatment, of men ≥65 years who have an average serum testosterone concentration < 275 ng/ and poor vitality, as defined by a score of <40 on the FACIT-Fatigue scale, will be associated with a greater percentage of men who have an improvement of ≥4 points in this test.

Secondary specific aims: To test the hypotheses that testosterone treatment for one year, compared with placebo, will result in more favorable outcomes in vitality/fatigue, as measured by the

1. SF-36 vitality scale,
2. Mood, as assessed by the Positive and Negative Affect Scales (PANAS)
3. PHQ-9 depression scale

Exploratory aims: Testosterone treatment for one year, compared with placebo treatment, will be associated with

1. A greater improvement in a patient global impression of vitality
2. A greater percentage of subjects who have an improvement of ≥4 units in the FACIT-Fatigue scale among men in all trials combined

2.4. Cognitive Function Trial

Primary specific aim: To test the hypothesis that testosterone treatment for one year, compared with placebo treatment, of men ≥65 years who have an average serum testosterone concentration < 275 ng/dL, who have subjective memory complaints as determined by their score on the MAC-Q questionnaire, and who demonstrate memory impairment as defined by a score on the Wechsler Memory Scale Revised Logical Memory II subscale recall (WMS-R LM II) or by Benton Visual Retention Test (BVRT) more than one SD below the performance for young men, aged 20-24 years [a criterion for age-associated memory impairment (AAMI) (40)] will result in greater improvement, or less decline, in verbal memory as assessed by the WMS-R LM II.

Secondary specific aims: To test the hypotheses that testosterone treatment for one year, compared with placebo, in the impaired subset of the study population defined above, will result in greater improvement or less decline in:

1. Visual memory assessed by the Benton Visual Retention Test (BVRT),
2. Spatial ability assessed by the Card Rotation test, and
3. Executive function/working memory assessed by the Trail Making Test (TMT).

Exploratory aims: Testosterone treatment for one year, compared with placebo treatment will result in greater improvement, or less decline in:

1. Verbal memory as assessed by the Wechsler Memory Scale Revised (WMS-R) Logical Memory II (WMS-R LM II) subtest, in all subjects, regardless of extent of memory impairment at baseline.The rationale for performing these tests in all subjects, regardless of presence or absence of impairment at baseline, is to determine if the cognitive response to testosterone depends on a demonstrated baseline impairment.
2. Patient global impression of change (PGIC) in memory,
3. Global cognitive function assessed by the Modified Mini-Mental State Examination (3MSE).

2.5. Anemia Trial

Primary Specific Aim: Testosterone treatment for a year, compared with placebo, of men who are anemic at baseline (hemoglobin concentration <13.5 g/dL) will be associated with a greater proportion whose anemia is corrected.

2.6. Measurements Across All Trials

The close coordination of the trials will permit measurements across all trials and hypotheses testing in the entire study group.

Secondary aim:

1. Testosterone treatment for one year, compared to placebo, of men in all trials will be associated with improved mood, as assessed by the Positive and Negative Affect Scales (PANAS) (41).

Exploratory aims:

1. Testosterone treatment for a year, compared with placebo, of men in all the trials, not just those who qualify for an individual trial, will be associated with better scores in each of the five primary end points.

2. Testosterone treatment for one year, compared with placebo, of all men in the study will be associated with a decrease in falls.
3. Testosterone treatment for one year, compared to placebo, of all men will be associated with better depression scores on PHQ-9.
4. Testosterone treatment for one year, compared with placebo, of men in all trials, will be associated with better scores on:
    - a patient global impression of change (PGIC) question in each primary efficacy area,
    - the sum of the PGIC questions in all primary efficacy areas, and
    - an overall PGIC question.
5. Testosterone treatment for one year, compared to placebo, will increase the incidence of a rise in prostate specific antigen (PSA), even after correction of the baseline value for low testosterone, sufficient to trigger a prostate biopsy.

3. Study Design

3.1. General Design

This study is designed as five separate, but highly coordinated, randomized, placebo-controlled clinical trials of the effect of testosterone in men ≥65 years who have a low serum testosterone concentration and symptoms and objective manifestations of abnormalities in the areas of physical function, sexual function, vitality and/or cognition.

The study will be conducted at 12 clinical sites across the United States. The data coordinating center at the University of Pennsylvania will coordinate the activities of the trial sites, central laboratory, central pharmacy, and associated reading centers. The trials are planned to take six years.

3.2. Study Endpoints

3.2.1. Physical Function Trial Endpoints

The physical function trial endpoints will be measured at 3, 6, 9 and 12 months.

Primary Endpoint:

- Mobility, as assessed by the 6-minute walk test

Secondary Endpoints:

- Physical function, as assessed by the physical function 10-item scale (PF10) of MOS SF36

Exploratory Endpoints:

- Patient global impression of change in walking a quarter mile
- Fall frequency in men in this trial and in all men

3.2.2. Sexual Function Trial Endpoints

The sexual function trial endpoints will be measured at 3, 6, 9 and 12 months.

Primary Endpoint:

- Overall sexual activity, as assessed by question 4 of the Harbor-UCLA 7-Day Sexual Function Questionnaire

Secondary Endpoints:

- Harbor-UCLA 7-day Sexual Function Questionnaire, Questions 1–3, and 5-6
- Libido, as assessed by the DISF-M-II
- Erectile function as assessed by International Index of Erectile Function (IIEF)

Exploratory Endpoints:

- Patient global impression of change in sexual activity

3.2.3. Vitality Trial Endpoints

The vitality trial endpoints will be measured at 3, 6, 9 and 12 months.

Primary Endpoint:

- Fatigue, as assessed by the 13-item FACIT-Fatigue Scale

Secondary Endpoints:

- Well-being, as assessed by the positive and negative scale (PANAS)
- Vitality scale of the SF-36
- PHQ-9 depression score

Exploratory Endpoints:

- Patient global impression of change in fatigue/vitality

3.2.4. Cognitive Function Trial Endpoints

The Cognitive Function Trial endpoints will be measured at 6 and 12 months. All subjects in all trials will be assessed by all cognitive function end points.

Primary Endpoint (in those subjects who have memory impairment at baseline):

- Verbal memory, as assessed by score on the WMS-R LM II

Secondary Endpoints (in those subjects who have memory impairment at baseline):

- Visual memory, as assessed by the BVRT
- Spatial ability, as assessed by the Card Rotations Test
- Working memory/executive function, as assessed by the Trail Making Test (B-A score)

Exploratory Endpoints:

- Patient global impression of change in cognitive function
- WMS-R LM II, BVRT, Card Rotations, and Trail Making Test in all subjects (both subjects who have memory impairment at baseline and those who do not)

3.2.5. Anemia Trial Endpoint

Primary Endpoint:

- Correction of anemia, as assessed by hemoglobin increasing from <13.5 to ≥13.5 g/dL.

3.3. Safety Measurements

Although this study will not have sufficient statistical power to assess the effect of testosterone on the safety parameters below, we shall monitor subjects for the development of these conditions because of the possibility that testosterone treatment could increase the risk.

3.3.1. Prostate Cancer

Prostate cancer will be diagnosed by prostate biopsy. Men will be referred for urologic evaluation for consideration of biopsy when either a prostate nodule is palpated on digital rectal examination or the serum PSA concentration increases ≥ 1.0 ng/mL above the testosterone-corrected baseline value, confirmed by a repeat determination.

3.3.2. Lower Urinary Tract Symptoms Due to Benign Prostatic Hyperplasia

Lower urinary tract symptoms will be evaluated by the International Prostate Symptoms Score (IPSS) questionnaire. An increase of >5 points or to an absolute value of >19 will result in a review of medications that affect urine flow rates and evaluation for prostatitis. If a cause is found, it should be treated. If no cause is found, treatment with an alpha blocker should be considered. If the subject is treated and symptoms persist, or if acute urinary retention occurs, the subject should be referred for urological consultation. If the urologist treats the subject and the score does not decrease below the above threshholds, gel treatment will be discontinued

3.3.3. Erythrocytosis

Erythrocytosis will be evaluated by hemoglobin. If the value increases to ≥ 17.5 g/dL, the subject will have repeat hemoglobin and testosterone measurements. If the testosterone concentration is above the target range, the number of depressions daily will be decreased and the hemoglobin repeated again. If the hemoglobin is still elevated, he will be referred for evaluation. If no treatable cause is found, the dose of testosterone will be decreased. At month 12, when treatment has stopped, men with elevated hemoglobin upon repeat will return for another hemoglobin test after 3 months time. The expectation is that after 3 months any effect the testosterone had on the hemoglobin will have dissipated.

3.3.4. Sleep Apnea

Men will be asked if they have been diagnosed with sleep apnea and are being treated. If they have been diagnosed but are not being treated, study medication will be discontinued. If subjects are being treated for sleep apnea, they will continue in the study.

3.3.5. Cardiovascular Disease

In order to determine if there is a relationship between testosterone treatment and cardiovascular events, we will administer a focused questionnaire about cardiovascular health at baseline and two others during treatment, one about incident cardiovascular events and one about incident symptoms. When a myocardial infarction, emergency revascularization, congestive heart failure, stroke or sudden death is reported, hospital records will be acquired

and evaluated. A committee of experienced cardiologists and neurologists will be appointed to adjudicate these events. The baseline questionnaire will allow us to assess balance between the two treatment arms in cardiovascular disease.The questionnaires about incident cardiovascular events and symptoms will help determine if testosterone treatment is associated with an increase in cardiovascular events. These questionnaires will be administered at each visit during the one year of treatment and also during the year of observation after treatment.

3.3.6. Fractures

Fractures will be monitored in follow up visits during the course of the trial. If a participant reports they had a fracture at a follow up visit the sites will follow up and request all possible medical records related to the fracture. This will include, but is not limited to, X-rays, CT scans, MRIs, other imaging exams, orthopedic and or operating notes. These documents will be sent for review to confirm the fracture diagnosis.

4. Subject Selection and Withdrawal

4.1. Number of Subjects

Subjects will be evaluated for study eligibility during Screening Visit 2. The total sample size is 800 for the entire study. Each clinical site is expected to enroll approximately 67 subjects. It is projected that 85% of subjects allocated to treatment will complete the 12 months of treatment.

4.2. Common Inclusion Criteria

The inclusion criteria common to all subjects in all trials are as follows:

- Men ≥65 years old
- Total serum testosterone concentration at screening visit 1 (SV1) <275 ng/dL, at screening visit 2 (SV2) < 300 ng/dL and an average serum testosterone concentration of < 275 ng/dL
- If the main T Trial has reached its enrollment goals, men must be eligible for either the Bone Trial or the CV Trial, if they are still open to enrollment (Please refer to the separate Bone Trial and CV Trial protocols for study details and study specific inclusion/exclusion criteria)

Blood should be collected from subjects who have been fasting (only water in the previous 8 hours). Only fasting samples are acceptable.

4.3. Common Exclusion Criteria

The exclusion criteria common to all subjects are as follows:

- Diagnosed prostate cancer or prostatic intraepithelial neoplasia (PIN) or, by the Prostate Cancer Risk Calculator, a >35% risk of having overall prostate cancer or >7% risk of having high grade prostate cancer
- Severe lower urinary tract symptoms (score of > 19) by the International Prostate Symptom Score questionnaire
- Hemoglobin <10 g/dL or >16.0 g/dL. Subjects who have hemoglobin level below 10 g/dL will be referred to their primary care providers for evaluation of anemia.

- Sleep apnea, diagnosed but untreated
- Alcohol or substance abuse within the past year (based on self report)
- Angina not controlled by treatment,
- NYHA class III or IV congestive heart failure
- Myocardial infarction within the previous 3 months
- Stroke within the previous 3 months
- Hypertension, defined as systolic blood pressure of >160 mm Hg or a diastolic blood pressure >100 mm Hg.
- Severe pulmonary disease that precludes physical function tests
- Serum creatinine >2.2 mg/dL; ALT 3x upper limit of normal; hemoglobin A1c >8.5%
- TSH > 7.5 mIU/L
- Kidney disease requiring dialysis
- Diagnosis or treatment for cancer within the past 3 years, with the exception of nonmelanotic skin cancers
- Body mass index (BMI) >37 kg/m^2
- Average testosterone concentration > 275 ng/dL; SV1 value > 275 ng/dL or an SV2 value of > 300 ng/dL
- Mini Mental State Exam (MMSE) Score <24

- Major psychiatric disorders, including major depression (PHQ-9 score > 14), mania, hypomania, psychosis, schizophrenia or schizoaffective disorders, that are untreated, unstable, have resulted in hospitalization or medication change within the previous three months, or would result in inability to complete the trial efficacy instruments. Subjects whose disorders have been stable while being treated for more than three months are eligible.
- Skin conditions at the testosterone gel application site, such as ulcer, erosion, lichenification, inflammation, or crust, or generalized skin conditions such as psoriasis or eczema that might affect testosterone absorption or tolerability of the testosterone gel
- Known skin intolerance to alcohol or allergy to any of the ingredients of testosterone gel

Medications:

Subjects who are using the following medications will be excluded:

- *Drugs that affect serum test*osterone concentration, (eg, testosterone, androstenedione, DHEA, estrogens, GnRH analogs, spironolactone, and ketoconazole) for 2 months during the previous 12 months or within the previous three months.
- rhGH or megesterol acetate within the previous three months.
- Anti-depressant medication that has been introduced within the past three months. (Subjects with diagnosed depression who have been stable for more than three months while taking anti-depressant medication are eligible.)
- Prednisone (dose of greater than 5 mg daily) use daily for more than two weeks, or equivalent doses of other glucocorticoids for more than two weeks during the previous three months.
- Opiate use within the past three months. Subjects who are using opiate analgesics intermittently for relief of chronic pain at doses that do not equal or exceed the equivalent of 20 mg methadone daily will be included. The following doses of opiate analgesics are considered equivalent:
    - Methadone 20 mg
    - Hydrocodone 30 mg

- Oxycodone 30 mg
- Morphine sulfate 30 mg
- Codeine sulfate 200 mg

4.3.1. Evaluation of T Level < 100

Men with a testosterone level < 100 ng/dL at SV1 or SV2 will be evaluated by the study physician or referred to an endocrinologist for the measures described below. Assessment of the following laboratory test results in combination will inform the physician of the need for further testing by MRI.

- serum LH > 9.3 mlU/mL
- total T4 < 4.5 µg/dL
- prolactin >30 ng/mL
- cortisol <10 µg/dL
- repeat testosterone

a. These five (5) tests will require a 10 cc venous blood draw. Blood must be drawn between 7 – 10 AM. Participants must be fasting for these tests which is defined as drinking only water after midnight of the night before the blood draw.

b. Men will be excluded who have a sellar mass >1 cm by an MRI scan of the head, in the absence of an elevated LH level.

c. Men will be excluded who have a history of mumps orchitis, castration, Klinefelter's syndrome or chemotherapy with an elevated LH level.

d. Clinical site staff must document that the participant has been told that standard medical treatment for a serum testosterone concentration < 100 ng/dL, is testosterone replacement, yet there is a 50% chance he would receive placebo for one year if he participates in The Testosterone Trial.

4.4. Inclusion and Exclusion Criteria for Physical Function Trial

Inclusion criteria: symptomatic mobility disability, defined by

- Self-reported difficulty in walking one-quarter mile and/or self-reported difficulty in walking up one flight of stairs and
- Walking speed <1.2 meters/second on the 6-min walk test

Exclusion criteria:

- Not ambulatory
- Other conditions affecting mobility of sufficient severity that testosterone is unlikely to improve, including neurological conditions (multiple sclerosis) and severe disabling arthritis of the lower extremity, joints, or back

4.5. Inclusion and Exclusion Criteria for Sexual Function Trial

Inclusion criteria:

- Self reported decreased libido and a sexual partner willing to have sexual intercourse ≥ twice/month
- Decreased libido, defined by a score of ≤20 on the DISF-M-II SR questionnaire

Exclusion criteria:

- Medical or nonmedical reasons that would preclude sexual activity (e.g., penile deformity, Peyronie's disease, pelvic surgery for bladder cancer)
- Severe peripheral vascular disease associated with an absence of pedal pulses
- Autonomic neuropathy

4.6. Inclusion and Exclusion Criteria for Vitality Trial

Inclusion Criteria:

- Decreased energy, self-reported
- Low vitality, defined by a score <40 on the FACIT-Fatigue Scale

4.7. Cognition Trial

Cognitive function tests will be performed in all men in all trials, so there will be no specific inclusion or exclusion criteria for this Trial

During the informed consent process, subjects will be asked for permission to audio-tape the testing sessions. Subjects may refuse and continue to participate in the study. This is done for quality control purposes at the Wake Forest University (WFU) Cognitive Function Reading Center. Recordings will be erased after scoring is completed.

4.8. Inclusion and Exclusion Criteria for Anemia Trial

Inclusion Criterion:

- Hemoglobin concentration <13.5 g/dL, the lower limit of normal for the central laboratory

Exclusion Criteria:

- Hemoglobin <10.0 g/dL

4.9. Subject Recruitment and Screening

The principal goals of recruitment are to identify men who have conditions that might be caused by a low testosterone concentration and who are representative of the United States population geographically, racially and ethnically. Recruitment techniques will include use of national media, local media, mass mailings by zip code, including retirement communities, retired employee groups (military, unions), graduates of local universities of appropriate graduating classes; local talks; direct recruitment at residential facilities for the elderly; focus groups to identify potential barriers to recruitment; and listing on ClinicalTrials.gov.

4.10. Early Withdrawal of Subjects

Because these trials are based on the principle of “intent-to-treat”, every attempt will be made to follow and evaluate all enrolled subjects for the duration of the trials. Therefore, even if treatment is discontinued, the subject will be asked to complete the appropriate evaluations.

5. Study Drug

5.1. Description

The study drug is AndroGel®(AbbVie, Inc., North Chicago, IL), which contains 1% testosterone in an alcohol-water gel and is FDA-approved for treatment of low testosterone in men. AbbViewill provide AndroGel in pumps, which deliver 1.25 g of gel per depression. AbbViewill also provide identical pumps with placebo gel.

5.2. Treatment Regimen

AndroGel or placebo will be applied to the abdomen, shoulders or upper arms once a day at the same time to dry, intact skin. Subjects will be instructed to wash their hands after application and to let the gel dry before dressing. It is important not to have contact with women or children while the gel is wet. They will also be asked not to bathe or get this area wet for five hours after application. Subjects will be taught how to apply the gel and they will be provided with written instructions and precautions. This information will be reviewed at each contact and visit.

The initial dose of AndroGel will be 5.0 g once a day. The serum testosterone concentration will be measured monthly for the first three months. If the testosterone concentration is not between 500 and 800 ng/dL at any time point, the dose will be either increased by increments of 1.25-2.5 g/day, up to a maximum of 15 g/day or decreased by increments of 1.25-3.75 ng/day. If the serum testosterone concentration is >800 ng/dL following two consecutive reductions in Androgel dose, treatment will be discontinued. A placebo subject will also be discontinued.

Men who stop treatment due to two consecutive reductions will have a repeat testosterone determination after two weeks. If the repeat testosterone value is <500 ng/dL (the lower limit of the target range) the participant will resume gel use. In this situation, the initial dose will be 1.25 g (one depression) a day, no matter how low the serum testosterone concentration. The matched placebo participant will resume gel use as well.

Men who are increased to the maximum dose of 15g/day will be asked to return for a serum testosterone determination within one month of the dose change. A subject from the placebo-treated group will also be asked to return for testosterone determination.

If the serum testosterone level is >1500 ng/dL the testosterone test will be repeated by the central lab, Quest. If the level is still >1500 ng/dL, the participant will be called in for an unscheduled blood draw for safety and so appropriate dosing can be insured. Levels of testosterone which are this high are typically caused by incorrect application or contamination at the blood draw site; for this reason proper gel application instructions will be reviewed at the time of the unscheduled blood draw. In order to maintain blinding of the sites, the DCC staff will instruct a site to also bring in a placebo subject for a repeat blood draw.

Furthermore in the case of a need for a repeat value due to a problem with the blood draw or an out of range value for any lab result, participants will be asked to return for a repeat blood draw and will be matched with a placebo participant to maintain the blind. The matching will be done by the DCC.

To maintain blinding when the dose of a subject in the testosterone group needs to be changed, a designated, unblinded DCC staff member will instruct the clinical trial site personnel to change that subject's dose, and also instruct that the dose (i.e., amount of gel) of a subject in the placebo group be changed at a randomly selected site (if possible) as well.

Reasonable efforts will be made to maintain blinding of investigators and staff members at clinical sites, provided such efforts do not impede subject safety.

5.3. Method for Assigning Subjects to Treatment Groups

Treatment assignment and balancing on prognostic factors will be done by the technique of minimization, rather than stratified randomization, because the sample size for this study (800) is not large enough to assure balance given the large number of strata that would be needed using the latter technique. Minimization will be performed by using a computer program developed at the Mayo Clinic in SAS Version 8. Factors for balancing for each of the five primary efficacy trials in which a subject may participate include study site, baseline serum testosterone concentration, age, and current use of an antidepressant. Additionally, use of a PDE-5 inhibitor will be balanced for those participating in the Sexual Function Trial.

5.4. Preparation and Administration of Study Drug

AndroGel pumps containing active and placebo gels will be supplied byAbbVieto the Investigational Drug Service (IDS) at the University of Pennsylvania, which will be the Central Pharmacy. The IDS will supply the pumps to the research pharmacies at each of the 12 trial sites. Subjects will be asked to return used pumps, which will be weighed. The weight will provide an assessment of the subject's compliance.

5.5. Storage

Bulk supplies of study medication will be stored in the central pharmacy at controlled room temperature (20-25 Celsius). Study medication that is labeled for individual study subjects and shipped to participating study sites will be stored at controlled room temperature (20-25 Celsius) with short temperature excursions allowed within the range of 15 to 30 Celsius.

5.6. Dispensing of Study Drug, and Return or Destruction of Study Drug

Blinded, tamper-sealed treatment kits containing a 3-month supply of testosterone or placebo, will be shipped to each site and stored securely. Each kit will be labeled with a specific randomization number, which will be repeated on each individually-labeled pump bottle. The initial set of pumps will be dispensed to the study subject only after randomization has taken place. Additional blinded and tamper-sealed sets of pumps will be provided to the sites in 3-month increments as refills, labeled for the individual study subject, after randomization has taken place. At appropriate intervals, there will be reconciliation of drug shipped, drug consumed, and drug remaining.

5.7. Concomitant Medications

Concomitant medications will be recorded. Subjects will be asked specifically if they are taking PDE5 inhibitors, antidepressants, antipsychotic drugs, or androgenic drugs.

6. Study Procedures and Visits

6.1. Telephone Prescreening

Potential subjects who call the trial site in response to advertisements or respond to a trial staff member at a health fair, etc., will be asked the following questions:

- Are you willing to answer questions about your possible participation in a testosterone research study?
- Are you 65 years of age or older?
- Do you have difficulty walking a quarter of a mile or climbing one flight of stairs?
- Has your desire for sex decreased?
- Is your energy is low?

Subjects will be asked several questions about major exclusion criteria such as recent use of testosterone, use of medications that affect bone, history of spinal surgeries and spinal conditions, history of cancer, stroke, heart attack, atrial fibrillation (if the CV Trial is open to enrollment), and height and weight to calculate body mass index. If a potential subject is willing to answer questions, is ≥65 years old, and not excluded by the medical history, he will be asked to schedule Screening Visit 1, the first in-person visit.

6.1.1. Screening Visit 1

Subjects will first be asked to give written, informed consent for Screening Visits 1 and 2 and to be assessed for eligibility for the Bone and/or CV Trial, using the Screening Consent Form.

Screening Visit 1 - Assessments and Procedures

- Screening Consent
- Brief medical and medications history
- Blood draw – 30 mL (Serum T, PSA reflex and chemistry panel reflex/ eGFR)

If the serum testosterone concentration is <275 ng/dL, and the risk of overall prostate cancer is ≤35% and of high grade prostate cancer ≤7%, as determined by the National Cancer Institute Prostate Cancer Risk Calculator, the subject will be asked to schedule Screening Visit 2.

If the subject has a testosterone level <100ng/dL at either screening visit 1 or screening visit 2, he will be evaluated as described in Section 4.3.1.

6.1.2. Screening Visit 2

The following procedures and questionnaires will be completed:

- Complete medical history, including medications
- Blood draw - 30 mL (Serum testosterone, CBC, Hgb A1c, TSH)
- Urinalysis
- Height and weight (for BMI); waist, hip and blood pressure measurements
- Digital rectal examination (DRE)
- International Prostate Symptom Score (IPSS)
- 6-Minute Walk Test (Physical Function Trial screening test)
- Derogatis Inventory of Sexual Function Male (Sexual Function Trial screening test)
- PHQ-9 (Trial eligibility depression screening test)
- MMSE (for exclusion of moderate to severe dementia)
- Interactive Voice Response (IVR) System instruction
- Please refer to the Bone Trial protocol and/or CV Trial protocol, for specifics of procedures that may need to occur at SV2 for those trials

Eligibility will be determined based on the results of these screening tests. Subjects who have a second testosterone concentration <300 ng/dL, and an average testosterone concentration between screening visit 1 and screening visit 2 of <275 ng/dL, meet all the common eligibility criteria, described in 4.4, and meet all of the inclusion and exclusion criteria for at least one of the Physical Function, Sexual Function or Vitality Trials, described in 4.5 - 4.8, will be asked to schedule a baseline visit.

6.1.2.1. Data Collection and Interactive Voice Response (IVRS) System

Several methods will be used to collect data from study subjects including self-administration and interviewer-completed questionnaires. Data from a few questionnaires will be collected using the Interactive Voice Response System. IVR is a computer-based, automated touch-tone telephone system used increasingly to collect self-reported, personally sensitive data.

Clinical site personnel will train subjects in the use of the IVR system during the second screening visit in preparation for data collection by IVR, prior to randomization. Subjects will be registered in the IVR system by their T Trial identification number. Each subject will be provided a secure username and password that they will be instructed to change the first time they access the IVR system. The subject will complete the FACIT Fatigue Scale during SV2 using the IVR system. If the T level indicates that a subject is eligible, subjects will be contacted by site personnel and instructed to submit the following forms via the IVR system before the baseline visit:

- UCLA 7-day diary
- PANAS
- PHQ-9.
- Baseline status questions: general health, physical function, sexual function, vitality and cognitive function.

Clinical site personnel will communicate with subjects regarding use of the IVRS and missed responses. Data from the IVR database will be transferred to the DCC database electronically.

6.1.3. Baseline Visit

The entire study, including rationale, assessments, treatment, and potential risks, will be described to the subjects who are deemed eligible. Subjects will be given the option of participating in one or more of the Physical Function, Sexual Function, or Vitality Trials, if they qualify for them. Those who agree will sign the Trial Informed Consent. All subjects will participate in the Cognitive Function Trial, and those who are anemic will be considered as participating in the Anemia Trial. Only subjects who qualify for and agree to participate in the Sexual Function or Vitality Trials will be tested by the secondary end points in that trial.

6.1.3.1. Assessments and procedures for all subjects

All subjects will be tested for the primary efficacy endpoints for all the trials and by other common endpoints as listed below.

- Concomitant medications
- International Prostate Symptom Score
- Cardiovascular History Questionnaire
- Weight (for BMI), waist, hip and blood pressure measurements
- Blood draw - 30 mL (serum testosterone, PSA, Hct/Hgb, creatinine, FSH and LH, extra serum archived for SHBG, DHT, estradiol; pharmacogenomics)

- Additional serum (60 mL), plasma (10 mL), and urine (10 mL) will be collected and stored at -80° for assay of 25 hydroxyvitamin D and unanticipated assays
- An additional 10 mL of blood will be drawn for development of lymphoblastoid lines from men who agree to sign the separate and optional Genetics Consent Form.
- Primary Efficacy Endpoints for Each Trial:
    - 6-Minute Walk Test (Physical Function Trial)
- Other Common Endpoints:
    - Patient Global Impression Questions
    - Falls
    - 3MSE, WMS-R LM II, BVRT, Card Rotations, and Trail Making Test
    - MAC-Q
    - SF-36 (entire form)
- Please refer to the Bone Trial protocol and/or CV Trial protocol for specifics of procedures that may need to occur at the Baseline Visit for those Trials

6.1.3.2. Secondary efficacy endpoints

Subjects will be tested for secondary efficacy endpoints only for those trials in which they are specifically enrolled, with the exception of the physical function trial secondary endpoints, which will be tested in all men. All endpoints in the Vitality Trial will be completed by IVR. In the Sexual Function Trial, the UCLA Questionnaire will be completed via IVR.

Baseline Visit – Secondary Efficacy Endpoints for Each Trial

- Subjects enrolled in the Physical Function Trial:
    - PF-10
- Subjects enrolled in the Sexual Function Trial:
    - DISF-M-II SR
    - IIEF
- Subjects enrolled in the Vitality Trial:
    - SF-36 Vitality scale (IVR)

6.1.3.3. Endpoint for Anemia Trial

All subjects will have blood drawn for hemoglobin and hematocrit at the baseline visit, and at 3, 6, 9 and 12 months. Subjects who are anemic at the baseline visit will considered to be enrolled in the Anemia Trial. They will require no additional tests.

6.1.3.4. Medication instructions

All subjects will be instructed in the use of the gel and given a three-month supply.

6.1.4. Months 1 and 2 Visits (± 7 days)

- Blood draw for serum testosterone (15 mL)
- Additional serum (20 mL), plasma (10 mL), and urine (10 mL) will be collected and stored at – 80°
- Concomitant medications
- Adverse Events
- Cardiovascular Event Questionnaire
- Cardiovascular Symptom Questionnaire
- Weigh used pumps
- Review gel application technique

- Dose adjustment, if necessary

After each of these visits, subjects will be notified by phone if an adjustment in gel dose is necessary.

6.1.5. Month 3 Visit (± 2 weeks)

6.1.5.1. Common assessments and procedures

- Concomitant medications
- Adverse events
- Cardiovascular Event Questionnaire
- Cardiovascular Symptom Questionnaire
- Weight, waist, hip and blood pressure measurements
- Digital rectal exam
- International Prostate Symptom Score
- Blood Draw – 30 mL (Serum T, PSA, Hct/Hgb; extra sera saved for SHBG, DHT, estradiol
- Additional serum (60 mL), plasma (10 mL), and urine (10 mL) will be collected and stored
- Primary Efficacy Endpoints for Each Trial:
    - 6-Minute Walk Test (Physical Function Trial)
    - UCLA Sexual Function Questionnaire question 4 (Sexual Function Trial, IVR)
    - FACIT-Fatigue (Vitality Trial, IVR)
- Other Common Endpoints:
    - Patient Global Impression of Change (PGIC) Questions (IVR)
    - Falls
    - Positive And Negative Affect Scales (IVR)
    - PHQ-9 (IVR)
- Weigh used pumps
- Review gel application technique
- Dose adjustment, if necessary
- Dispense medication for three months

6.1.5.2. Secondary and exploratory efficacy endpoints for each trial.

Secondary endpoints will be performed only on subjects specifically enrolled in that trial, with the exception of the physical function trial secondary endpoints, which will be tested in all men enrolled in the Trial.

<u>Month 3 Visit – Secondary and Exploratory Efficacy Endpoints for Each Trial</u>

- Physical Function
    - PF-10
    - PGIC question for physical function (IVR)
- Sexual Function
    - UCLA Sexual Function Questionnaire-complete (IVR)
    - DISF-M-II SR
    - IIEF
    - PGIC question about sexual function (IVR)
- Vitality
    - SF-36 Vitality scale (IVR)

- PGIC question about vitality/fatigue (IVR)

6.1.6. Months 4 and 5 Assessments (± 7 days)

Subjects will be asked by **telephone** about adverse events, concomitant medications, and gel use.

- Concomitant medications
- Adverse events
- Review of instructions for use of testosterone gel

6.1.7. Month 6 Visit (± 2 weeks)

The Month 6 visit assessments will be similar to those of the Month 3 visit.

- Concomitant medications
- Adverse events
- Cardiovascular Event Questionnaire
- Cardiovascular Symptom Questionnaire
- Weight, waist, hip and blood pressure measurements
- Blood Draw – 30 mL (Serum T, Hct/Hgb; extra sera saved)
- Additional serum (60 mL), plasma (10 mL), and urine (10 mL) will be collected and stored at – 80°
- Primary Efficacy Endpoints for Each Trial:
  - 6-Minute Walk Test (Physical Function Trial)
  - UCLA Sexual Function Questionnaire - question 4 (Sexual Function Trial, IVR)
  - FACIT-Fatigue (Vitality Trial, IVR)
- Other Common Endpoints:
  - Patient Global Impression of Change Questions (IVR)
  - Falls
  - Positive And Negative Affect Scales (IVR)
  - WMS-R LM II, BVRT, Card Rotations, and Trail Making Test
  - MAC-Q
  - PHQ-9 (IVR)
- Weigh used pumps
- Review gel application technique
- Dose adjustment, if necessary
- Dispense medication for three months

<u>Month 6 Visit – Secondary and Exploratory Efficacy Endpoints for Each Trial</u>

- Physical Function
  - PF-10
  - PGIC question for physical function (IVR)
- Sexual Function
  - UCLA Sexual Function Questionnaire-complete (IVR)
  - IIEF
  - PGIC question about sexual function (IVR)
- Vitality
  - SF-36 Vitality scale (IVR)
  - PGIC question about vitality (IVR)

6.1.8. Months 7 and 8 Assessments (± 7 days)

Subjects will be asked by telephone about adverse events, concomitant medications and gel use.

- Concomitant medications
- Adverse events
- Review of instructions for use of testosterone gel

6.1.9. Month 9 Visit (± 2 weeks)

Assessments and procedures at the Month 9 visit will be similar to those at the Month 3 visit except that prostate evaluation will not be performed.

- Concomitant medications
- Adverse events
- Cardiovascular Event Questionnaire
- Cardiovascular Symptom Questionnaire
- Weight, waist, hip and blood pressure measurements
- Blood Draw – 30 mL (Serum T, Hct/Hgb; extra sera saved)
- Additional serum (60 mL), plasma (10 mL), and urine (10 mL) will be collected and stored at – 80°
- Primary Efficacy Endpoints for Each Trial:
    - 6-Minute Walk Test (Physical Function Trial)
    - UCLA Sexual Function Questionnaire question 4 (Sexual Function Trial, IVR)
    - FACIT-Fatigue (Vitality Trial, IVR)
- Other Common Endpoints:
    - Patient Global Impression of Change Questions (IVR)
    - Falls
    - Positive And Negative Affect Scales (IVR)
    - PHQ-9 (IVR)
- Weigh used pumps
- Review gel application technique
- Dose adjustment, if necessary
- Dispense medication for three months

<u>Month 9 Visit – Secondary and Exploratory Efficacy Endpoints for Each Trial</u>

- Physical Function
    - PF-10
    - PGIC about physical function (IVR)
- Sexual Function
    - UCLA Sexual Function Questionnaire-complete (IVR)
    - IIEF
    - PGIC about sexual function (IVR)
- Vitality
    - SF-36 Vitality scale (IVR)
    - PGIC about vitality (IVR)

6.1.10. Months 10 and 11 Assessments (± 7 days)

Subjects will be asked by telephone about adverse events, concomitant medications, and gel use.

- Concomitant medications
- Adverse events
- Review of instructions for use of testosterone gel

6.1.11. Month 12 Visit (± 2 weeks)

Month 12 will be the end of treatment. All common and trial-specific assessments will be made at this visit for all trials.

Common Assessments and Procedures

- Concomitant medications
- Adverse events
- Cardiovascular Event Questionnaire
- Cardiovascular Symptom Questionnaire
- International Prostate Symptom Score
- Height, weight, waist, hip and blood pressure measurements
- Blood Draw – 30 mL (Serum T, PSA, Hct/Hgb, HgbA1c, chemistry panel; extra sera saved for SHBG, DHT, estradiol)
- Additional serum (60 mL), plasma (10 mL), and urine (10 mL) will be collected and stored
- Digital rectal exam
- Primary Efficacy Endpoints for Each Trial:
    - 6-Minute Walk Test (Physical Function Trial)
    - UCLA Sexual Function Questionnaire question 4 (Sexual Function Trial, IVR)
    - FACIT-Fatigue (Vitality Trial, IVR)
- Other Common Endpoints:
    - Patient Global Impression of Change Questions (IVR)
    - Falls
    - Positive And Negative Affect Scales (IVR)
    - 3MSE, WMS-R LM II, BVRT, Card Rotations, and Trail Making Test
    - MAC-Q
    - PHQ-9 (IVR)
- Weigh used pumps
- Please refer to the Bone Trial protocol and /or CV Trial protocol for specifics of procedures that may need to occur at the Month 12 Visit for these trials

Month 12 Visit – Secondary and Exploratory Efficacy Endpoints for Each Trial

- Physical Function
    - PF-10
    - PGIC about physical function (IVR)
- Sexual Function
    - UCLA Sexual Function Questionnaire-complete (IVR)
    - DISF-M-II SR
    - IIEF
    - PGIC about sexual function (IVR)
- Vitality
    - SF-36 Vitality scale (IVR)
    - PGIC about vitality (IVR)

6.1.12. Months 18 and 24 Assessments (± 1 month)

These are post-treatment assessments. Month 18 visit will occur at the trial site. The Month 24 visit will be conducted over the telephone.

Months 18 and 24 Assessments

- Blood draw – 15 mL for serum PSA – Month 18 only
- Adverse events
- Cardiovascular Event Questionnaire
- Cardiovascular Symptom Questionnaire

6.2. Subject Compensation

Subjects will be compensated during the course of the trial, based on the number of visits completed and the number of trials in which they participate. In addition, travel and parking expenses, and meal tickets will be provided for study visits.

6.3. Screening, Assessment, & Monitoring Schedule

PROCEDURES	Phone	Screen Visit 1	Screen Visit 2	Baseline	Treatment Month												Post-Rx	
					1	2	3	4	5	6	7	8	9	10	11	12	18	24
Screening & Monitoring																		
Symptom screen	X																	
Serum T		X	X	X	X	X	X			X			X			X		
Efficacy trial screening			X															
PSA		X		X			X									X	X	
IPSS			X	X			X									X		
Urine sample stored			X	X	X	X	X			X			X			X		
Hct/Hgb			X	X			X			X			X			X		
Hx for clinical events	X			X			X			X			X			X	X	X
Medical history		X	X															
DRE			X				X									X		
Cardiovascular Hx				X														
CV Events and Sx					X	X	X			X			X			X	X	X
Primary Efficacy Endpoints																		
Phys Fx				X			X			X			X			X		
Sexual Fx				X			X			X			X			X		
Vitality				X			X			X			X			X		
Cognition				X						X						X		
Secondary Efficacy Endpoints																		
Phys Fx				X			X			X			X			X		
Sexual Fx				X			X			X			X			X		
Vitality				X			X			X			X			X		
Cognition				X						X						X		
Other Efficacy Area																		
Anemia				X			X			X			X			X		
Measures Across Trials																		
PGICs				X			X			X			X			X		
Falls				X			X			X			X			X		
PANAS				X			X			X			X			X		
PHQ-9				X			X			X			X			X		

Abbreviation: IPSS, International Prostate Symptom Score; PGIC, Patient Global Impression of Change questions; PANAS, Positive and Negative Affect Scales

Primary efficacy endpoints will be assessed in all subjects; secondary efficacy endpoints will be assessed only in those who specifically qualify for that trial, except the endpoints for the Cognition Trial, which will be assessed in all subjects.

7. Statistical Plan

7.1. Analytical Methods and Sample Size Estimations: Overview

7.1.1. Analysis of Primary and Secondary Endpoints for the Individual Trials

Each of the efficacy trials (Physical Function, Sexual Function, Vitality, Cognitive Function and Anemia) is considered a separate trial, so the results will be analyzed separately. The primary and secondary endpoints of each of these trials will be evaluated for those subjects who participate in each specific trial. Primary analysis of outcomes with interim measures in addition to baseline and 12 months measures will be performed with random effects models for longitudinal data. Logistic models will be used for binary variables. Outcomes with measures at baseline and 12 months only will be compared using chi-square tests (binary outcomes) or Student's T test (continuous variables). Wilcoxon's rank sum test will be used for continuous variables that deviate substantially from a normal distribution. Dichotomous outcomes have been selected rather than continuous ones in order to determine not only if testosterone has a statistically significant effect compared to placebo, but if it also has an effect that is of clinical significance. All analyses will be adjusted for balancing factors. We shall perform sensitivity analyses to assess the potential impact of missing data by fitting shared parameter models that relax the missing at random assumptions of the proposed random effects models. We shall employ the methods of Benjamini and Hochberg to control for the impact of multiple analyses. Individuals will be analyzed in the group to which they were randomized regardless of compliance with assigned treatment (intention-to-treat principle), but sensitivity analyses accounting for compliance will be performed.

7.1.2. Sample Size Estimation

Sample sizes for each efficacy area were calculated based on two-sided 0.05 level tests and 90% power. Although our primary analyses will include data from each visit, we have calculated sample sizes based on tests considering only values at baseline and 12 months, which provide conservative estimates of sample size; therefore, we apply only a modest inflation factor of 5% to help compensate for subjects who drop out early.

We do not expect that all trials will complete enrollment at the same time. Once a trial reaches its target enrollment, no further subjects will be enrolled in that trial unless they qualify for that trial and a trial that remains open. Open Trials include the Bone Trial and CV Trial in combination of one of the 3 main trials (sexual, physical or vitality).

7.1.3. Physical Function Trial

We estimated the expected changes in 6-minute walking distance on the basis of unpublished data from the control group in the Walking and Leg Circulation Study (WALCS), in which the subjects performed the 6-minute walk at baseline and after one year. In this study the proportion of untreated subjects with an increase of 50 m or more at 12 months was 16%. To detect an increase from 15% to 30% with 90% power will require 350 subjects, inflated to 370 to compensate for dropout before the three-month visit.

We shall compare the proportions in each treatment group who achieve an increase of ≥ 8 points on the PF10 because such an improvement has been shown to be clinically meaningful.

We shall also compare the actual distributions of changes in distance on the 6-minute walk, and in PF10 scores, and differences in proportions with a 50m or greater decline in distance covered during the 6-minute walk, and in proportions experiencing one or more falls.

7.1.4. Sexual Function Trial

Published data suggest that testosterone treatment increases the mean sexual activity score (question 4 in the UCLA 7-day diary) by 0.75 units (SD of change: 1.86) (42). This difference of 0.75 units also appears to be clinically meaningful, in that hypogonadal men treated with a testosterone gel who increased their score by at least this amount had the same distribution of scores as eugonadal men. A sample size of 262 subjects (275 to compensate for missed visits) will be needed to detect this difference with high power.

7.1.5. Vitality/Fatigue

A change of 4 points on the FACIT-Fatigue Scale has been demonstrated to be a clinically meaningful difference (43, 44). We shall compare the proportions experiencing such a change in the two treatment groups. Because self-reported outcomes often show a substantial placebo response rate, we assume that 20% of those receiving placebo will show an improvement of 4 or more points on the FACIT-Fatigue Scale. A sample size of 420 will provide 90% power to detect an increase in this proportion to 35% in the testosterone arm.

7.1.6. Cognition Trial

We aim to detect an effect size of 0.3 (based on change from baseline to 12 months), which corresponds to a 3-point improvement in Paragraph Recall. On the WMS-R Logical Memory II Subscale Recall, the scaled-score that corresponds to the 50th percentile performance for a man 70-74 years old is 17; an effect size of 0.3, or 3 point increase, would improve that score to 20, corresponding to the 50th percentile performance for a man 45-54 years old. These data suggest that a 3-point difference will be clinically significant. The sample size required to attain 90% power for this difference is 235 per arm, 470 for both arms, or 500 to help compensate for missed visits. Based on previous studies, we expect that approximately 60% of men enrolled will have memory impairment more than one SD below the performance for young men, aged 20-24 years, a criterion for age-associated memory impairment. Therefore, of the 800 subjects enrolled in the Core Testosterone Trial, we expect that approximately 480 men, expected to be evenly distributed between treatment arms, will demonstrate age-associated cognitive impairment at baseline, as defined above.

7.1.7. Anemia Trial

We shall identify all subjects who have low hemoglobin at baseline and compare the proportions who are no longer anemic over the 12 months of treatment. We expect 10-20% of subjects to be anemic at baseline, providing 80 – 160 subjects in whom we shall evaluate the effect of testosterone on anemia. Assuming 10% of those assigned to placebo become non-anemic, we shall have 80-90% power to detect improvements ranging from 15 to 26 percentage points depending on the baseline proportion anemic.

7.2. Analytic Plans for Measures Across All Trials

7.2.1. Efficacy Endpoints from Individual Trials

The primary efficacy endpoint for each trial will be evaluated in all subjects, but the results will be considered exploratory. These analyses will take into account whether or not the subject

had actually participated in the trial associated with a given endpoint, in addition to all baseline balancing factors. Similarly, secondary endpoints from the physical function trial (falls) and the vitality trial (PANAS) will be evaluated in all subjects.

7.2.2. Patient Global Impression of Change (PGIC)

For the Physical and Sexual Function, Vitality and Cognition Trials and for global assessment overall, a seven-point Likert-scale for PGIC will be administered every three months. These data will be evaluated at each time point and over the entire 12-month observation period. In addition to a score for each trial for subjects who specifically participated in a trial and another for all subjects in all trials together, we will sum all scores to generate an overall score. There will also be a score for the overall questionnaire. In addition, we shall evaluate the extent to which the Likert scale outcomes are consistent with the changes in objective measures for subjects in each trial.

7.3. Adverse Events

We will compare proportions of men experiencing adverse events in each treatment group, with particular attention to areas that are plausibly associated with testosterone, including erythrocytosis, urinary tract symptoms, and prostate-related events.

7.3.1. Prostate Cancer

This is the safety parameter of primary interest and the focus of our interim monitoring plan described in Section 7.5. In addition to monitoring diagnosis of cancer, we will calculate rates and confidence intervals for biopsy requirement, and grade of cancer in those with cancer diagnoses.

7.4. Sample Size for the Entire Study

The total sample size for all trials is 1051, as shown in the table below. Assuming that approximately 33% of these men will qualify for, and participate in, two efficacy areas, the sample size for the entire study becomes 800 (1051 x 3/4 = 788, rounded to 800).

Table 7.4. Sample Sizes for Individual Trials and All Trials

Trial	One Arm	Both Arms	+5%
Physical function	174	348	366
Sexual function	131	262	275
Vitality	200	400	420
All Trials	505	1010	1051

The assumption that 33% of men will qualify for at least two trials comes from unpublished data in two studies. One is the Rancho Bernardo Study, showing that in response to the ADAM (Androgen Deficiency in the Aging Male) questionnaire of symptoms in three areas (energy, strength, sexual), 36% of men had symptoms in at least two areas, and some in three (unpublished). Because the men in the proposed study will all have low testosterone concentrations, the overlap may be even higher, an assumption supported by data from men in the EMAS study who were >65 years old and had testosterone concentrations <250 ng/dL. We now estimate that at least 33% of subjects will participate in two trials and 10% in three, but we

base our sample size estimates on the conservative assumption of only 33% participating in two trials.

We shall allow a variable degree of over enrollment in the main trials; Physical Function Trial, Sexual Function Trial and Vitality Trial if necessary to complete enrollment in one of the other trials, eg Bone Trial & CV Trial.

7.5. Interim monitoring

Interim monitoring in this trial will focus on safety; there is no intent to consider early stopping on the basis of any efficacy parameter. The primary safety concern related to testosterone treatment is increased risk of prostate cancer. Evaluating this risk during the study in an accurate and unbiased manner will not be possible, for several reasons. Approximately 60% of men this age harbor occult prostate cancer, and even after we select men who have reduced risk, we expect as many as 20% of the subjects will have biopsy-detectable cancer at study entry (unpublished data from the PCPT, rate for men >65 years). Thus, for any biopsy performed as a result of PSA changes or DRE finding, the probability of a positive finding will be at least 20%, yielding as many as 80 cases of prostate cancer per treatment arm. Because testosterone is known to cause PSA to rise, we might expect to perform more biopsies in the testosterone-treated group, and therefore might diagnose more cancers in that group, whether or not testosterone actually increases prostate cancer risk, i.e., we might have ascertainment bias. Further, the PSA increases in men receiving testosterone might be selectively observed in men with occult cancer, because testosterone may “unmask” such cancers, whether or not it exacerbates their growth. Therefore, even a large difference in numbers of cancer diagnoses between arms might not necessarily indicate a difference in cancer risk. On the other hand, any diagnosis of prostate cancer may lead to cancer treatment, which has its own potential risk of major adverse effects, particularly on quality of life. If testosterone truly does not increase cancer risk, but does increase risk of diagnosis of indolent tumors that are likely to remain asymptomatic during a man’s lifetime, then these diagnoses in themselves represent an adverse consequence of treatment. By adjusting the PSA for serum testosterone, however, we might mitigate the possibility of ascertainment bias.

Given these considerations, which impose great difficulties on the development of a statistical monitoring plan, we propose to use an approach that balances benefits and risks. We assume a rate of cancer diagnosis in the placebo arm of 1%/year, based on unpublished data from the PCPT, and a follow-up time of 24 months for each subject. Under these assumptions, we expect a total of 8 cases per arm under the null hypothesis of no excess cancers in testosterone-treated subjects. We propose as a basis for monitoring cancer diagnosis a one-sided O’Brien-Fleming boundary with an overall alpha of 0.20 and with a Lan-DeMets spending function modification for comparing time to cancer diagnosis. This plan provides 90% power for detecting a hazard ratio of 2.4 or higher. We specify a looser criterion for early stopping than would be typical for an efficacy boundary while still maintaining the probability of error at a relatively low level. We shall perform three interim analyses, specifically, after 25%, 50% and 75% of our target sample size has completed 12 months of follow up. Use of the spending function approach will permit additional analyses or a modified schedule, should the DSMB so request.

8. Safety and Adverse Events

8.1. Definitions

Definitions are per the January 2007 Guidance on Reviewing and Reporting Unanticipated Problems Involving Risks to Subjects or Others and Adverse Events, Office on Human

Research Protection (OHRP) Guidance. http://www.hhs.gov/ohrp/policy/AdvEvntGuid.htm. The requirements and processes for reporting adverse events are described in the corresponding National Institute on Aging (NIA) Guidelines.

8.1.1. Adverse Event

An *adverse event (AE)* is any untoward or unfavorable medical occurrence in a human study participant, including any abnormal sign (for example, abnormal physical exam or laboratory finding), symptom, or disease, temporally associated with the participant's involvement in the research, whether or not considered related to the subject's participation in the research.

8.1.2. Serious Adverse Event

A *serious adverse event* (SAE*)* is any AE that is:

- fatal
- life-threatening
- requires or prolongs hospital stay
- results in persistent or significant disability or incapacity
- results in congenital anomalies or birth defects
- an important medical event*

Important medical events* are those that may not be immediately life threatening, but are clearly of major clinical significance.

8.1.3. Unanticipated Problem

An Unanticipated Problem is any incident, experience, or outcome that meets all of the following criteria:

- unexpected (in terms of nature, severity, or frequency) given the research procedures that are described in the IRB-approved research protocol and informed consent document;
- related or possibly related to participation in the research; possible related means that there is a reasonable possibility that the incident, experience or outcome may have been caused by the procedures involved in the research.
- suggests that the research places subjects or others at a greater risk of harm (including physical, psychological, economic, or social harm) related to the research than was previously known or recognized.

8.1.4. Adverse Event Reporting Period

The study period during which adverse events must be reported is normally defined as the period from the initiation of any study procedures to the end of the study treatment follow-up.

8.1.5. Preexisting Condition

A preexisting condition is one that is present at the time of signing the consent form for the main study. A preexisting condition should be recorded as an adverse event if the frequency, intensity, or the character of the condition worsens during the study period.

8.1.6. General Physical Examination Findings

At screening, any clinically significant abnormality should be recorded as a preexisting condition. At the end of the study, any new clinically significant findings/abnormalities that meet the definition of an adverse event must also be recorded and documented as an adverse event.

8.1.7. Post-study Adverse Event

All unresolved adverse events should be followed by the investigator until the events are resolved, the subject is lost to follow-up, or the adverse event is otherwise explained. At the last scheduled visit, the investigator should instruct each subject to report any subsequent event(s) that the subject, or the subject's personal physician, believes might reasonably be related to participation in this study. The investigator should notify the study DCC of any death or adverse event occurring during the year after a subject has completed treatment.

8.1.8. Abnormal Laboratory Values

A clinical laboratory abnormality should be documented as an adverse event if any one of the following conditions is met:

- The laboratory abnormality is not otherwise refuted by a repeat test to confirm the abnormality
- The abnormality suggests a disease and/or organ toxicity
- The abnormality is of a degree that requires active management

8.1.9. Hospitalization, Prolonged Hospitalization or Surgery

Any adverse event that results in hospitalization should be documented and reported as a serious adverse event. Any condition responsible for surgery should be documented as an adverse event if the condition meets the criteria for an adverse event and reported as a severe adverse event if hospitalization is required. Neither the condition, hospitalization, nor surgery is reported as an adverse event if the hospitalization was for diagnostic or elective surgical procedures for a preexisting condition.

8.2. Recording of Adverse Events

At each contact with the subject, the investigator must seek information on adverse events by specific questioning and, as appropriate, by examination. Information on all adverse events should be recorded immediately in the source document, and also in the appropriate adverse event module of the case report form (CRF). All clearly related signs, symptoms, and abnormal diagnostic procedures results should be recorded in the source document. All adverse events occurring during the study period must be recorded. The clinical course of each event should be followed until resolution, stabilization, or until it has been determined that the study treatment or participation is not the cause. Serious adverse events that are still ongoing at the end of the study period must be followed up to determine the final outcome. Any serious adverse event that occurs during the year after completion of treatment will similarly be recorded and reported.

8.3. Reporting of Serious Adverse Events

8.3.1. Study Sponsor Notification by Investigator

Clinical sites are required to report serious adverse events to the DCC, within 24 hours of first knowledge of the event. The DCC will facilitate the timely reporting and updates to regulatory authorities, the DSMB, NIH and the FDA according to the standard MedWatch guidelines. A

Serious Adverse Event (SAE) form must be completed by the investigator and faxed to the DCC within 24 hours. The DCC will report all SAEs to the DSMB Chairman and DSMB safety Monitor within 48 hours of first knowledge of the event. The investigator will keep a copy of this SAE form on file at the study site. At the time of the initial report, the following information should be provided:

- Study identifier
- Study Center
- Subject number
- A description of the event
- Date of onset
- Current status
- Whether study treatment was discontinued
- The reason why the event is classified as serious
- Investigator assessment of the association between the event and study treatment

Within the following 48 hours, the investigator must provide further information on the serious adverse event in the form of a written narrative. This should include a copy of the completed Serious Adverse Event form, and any other diagnostic information that will assist the understanding of the event. Significant new information on all ongoing serious adverse events should be provided promptly to the study sponsor.

8.3.2. IRB Notification by Investigator

Reports of all serious adverse events (including follow-up information) must be submitted to the IRB within 10 working days. Copies of each report and documentation of IRB notification and receipt will be kept in the Clinical Investigator's binder.

8.4. Unblinding

Treatment assignment will be blinded to all but a single designated "unblinded" physician at the trial site. Although testosterone treatment might increase the risk of certain diseases, such as prostate cancer, lower urinary tract symptoms due to benign prostatic hyperplasia, or erythrocytosis, the blind will not be broken even if a subject develops one of these conditions during the study. Instead, the following approach will be taken.

a. If a subject is diagnosed with prostate cancer during the study, treatment will be discontinued, whether the treatment is testosterone or placebo.

If the subject's score on the International Prostate Symptoms Score increases by > 5 points above the baseline value or to an absolute score of >19, suggesting worsening of lower urinary tract symptoms, the subject will be referred to a urologist for evaluation of medications that affect urine flow rate and for prostatitis. Treatment with an alpha blocker will be considered. If the subject's score does not decrease below the above threshold in response to treatment, gel treatment will be discontinued.

b. If a subject develops a hemoglobin > or = 17.5 g/dL, he will be evaluated for causes of secondary erythrocytosis. If none are found, the dose of gel will be lowered. If the hemoglobin is still > or = 17.5 g/dL, treatment will be discontinued.

8.5. Stopping Boundaries

Table 8.5. Stopping Boundaries

Interim Analysis	Nominal Alpha for Boundary	Hazard Ratio on Boundary
1	0.01	4.9

2	0.067	2.1
3	0.118	1.6
Final	0.157	1.4

Because of the considerations described in Section 7.5, Interim Monitoring, we would not want to base early stopping solely on the cancer cases, as these may be subject to substantial bias — any of the boundary scenarios outlined in Section 7.5. are possible without any true excess risk due to the likelihood of ascertainment bias. We shall ask the DSMB to consider the cancer data together with the interim efficacy data from all the trials. Should the interim data suggest no emerging benefit, the stopping boundary shown might be applied. If the interim data are consistent with potentially valuable effects of treatment, however, a somewhat greater imbalance in cancer cases might be tolerated. The proportion of cases on each arm that are of high grade (Gleason score ≥7) will also be a consideration, but the number of such cases we expect to observe in this trial will be small, perhaps a fourth of all cases. We shall also ask the DSMB to consider the extent to which ascertainment bias might affect the comparison of cancer rates in each arm, and in this regard shall present relevant data, eg the number of biopsies by arm and the proportion of cancers among those with biopsies.

8.6. Monitoring Subject Safety

8.6.1. Potential Risks to Subjects

Several conditions to which elderly men are particularly prone are, at least partly, testosterone-dependent. These and other potential risks are described below:

8.6.1.1. Prostate cancer.

The basis for monitoring men in a testosterone trial for prostate cancer is that it is, to some degree, testosterone-dependent, and because elderly men often harbor clinically silent prostate cancer. There is no direct evidence, however, that either endogenous serum testosterone concentrations or testosterone treatment of men with low testosterone concentrations increases the risk of clinical prostate cancer.

8.6.1.2. Prostate biopsy.

Prostate biopsy will be performed if medically indicated. The two primary risks of this biopsy, which is performed by a transrectal, ultrasound guided approach, are bleeding and infection. By taking proper precautions, the risk of these complications requiring hospitalization is <1%.

8.6.1.3. Benign prostatic hyperplasia.

Testosterone treatment of elderly hypogonadal men might also increase the risk of lower urinary tract symptoms, because the prostate is a testosterone-dependent gland and because BPH is common in these men.

8.6.1.4. Erythrocytosis.

One potential consequence of testosterone treatment is the development of erythrocytosis. We shall therefore evaluate the men who participate in this study to determine if those whose hemoglobin values are normal before treatment experience an increase above normal (erythrocytosis) during treatment.

8.6.1.5. Sleep apnea.

Another potential risk is exacerbation of sleep apnea, although the evidence is weak.

8.6.1.6. Physical function testing.

There is a very small risk of injury from a fall or ankle sprain during the 6-minute walk test.

8.6.1.7. Sexual function and vitality testing.

The potential risks of these studies are the time the testing takes, minor distress of answering questions of a personal nature, and the fear of lack of confidentiality.

8.6.1.8. Time burden.

The large number of tests proposed could be tiring for an elderly man, especially one who participates in more than one protocol. In a pilot study in which 10 men, mean age 75 years, at each of three sites (30 men total) were administered all of the tests for all trials, the mean time for completion of all the tests was <100 min, and the subjects found most of the tests relatively easy. However, there was variability, and a few subjects took longer. Trial site personnel should be cognizant that some subjects could have difficulty in participating in multiple trials.

8.6.2. Protection Against Risk

Subject selection and monitoring procedures have been designed to minimize the risks. First, we shall select subjects who are at low risk of the potential side effects. Second, we shall employ procedures to minimize the potential risks. Third, we shall monitor enrolled subjects for the potential side effects.

8.6.2.1. Erythrocytosis.

A potential subject will be enrolled only if his hemoglobin is ≤16 g/dL. Men who enroll will be monitored by hemoglobin at 3, 6, and 12 months. An increase above the upper limit of normal (17.5 g/dL) in either treatment group will lead first to repeat measurements of hemoglobin and testosterone. If the serum testosterone level is above the target range, , the gel dose will be decreased. If the repeat hemoglobin is still elevated, the subject will be referred for evaluation for causes of erythrocytosis and, if found, treatment. If no cause of secondary erythrocytosis is found, or, if erythrocytosis does not return to.

normal within one month, treatment will be discontinuedThe external (unblinded) physician who evaluates subjects for erythrocytosis will consider all standard treatments including phlebotomy.

The exception to this is month 12, at which time all men stop treatment/placebo. At the month 12 visit, if a man has an elevated hemoglobin upon repeat, he will be brought back in after 3 months of being off of treatment. At which point it is expected his hemoglobin will have lowered. If the hemoglobin has not been lowered after 3 months, men will be referred.

8.6.2.2. Prostate cancer.

We shall exclude men with diagnosed prostate cancer or prostatic intraepithelial neoplasia (PIN). Men will also be excluded who have a >35% risk of having a prostate cancer and a >7% risk of having high grade prostate cancer by the Prostate Cancer Risk Calculator (http://www compass.fhcrc.org/edrnnci/bin/calculator/main.asp). This Risk Calculator will be used because it takes into account not only PSA, but also other known risk factors, including age, race, family

history, and previous biopsy and is therefore more conservative and exposes the subjects to less risk than if exclusion were based only on PSA.

Risk will be reduced further by adjusting the baseline PSA concentrations upward to account for the likelihood that those concentrations are lower than they would have been had the subjects' testosterone concentrations been normal. Each man's PSA will be adjusted to what would be expected if his serum testosterone were 460 ng/dL. The adjusted PSA would then be used in the Prostate Risk Calculator. Although adjusting the PSA for serum testosterone is not standard clinical practice, we think this approach is preferable to using the unadjusted value because it takes into account the physiologic relationship between testosterone and PSA and because, by raising the PSA, it is more conservative.

The use of the Risk Calculator, instead of PSA alone, allows us to account not only for PSA, but also other known risk factors, including age, race, family history, and previous biopsy and therefore is more conservative and exposes subjects to less risk than if exclusion were based only on PSA. It illustrates to a subject that every man of this age has some risk.

The rationale for choosing 35% of overall prostate cancer risk and 7% risk of high grade cancer is two-fold: 1) It is low enough to be quite conservative. For example, for a 65 or 75 year-old white man with no other risk factors to have a risk of ≤35%, his PSA would need to be ≤3.0 ng/mL, which would not be a cause for biopsy in routine clinical care. 2) It is high enough to include enough subjects that recruitment will still be practical.

Subjects will be monitored during the one year of treatment by repeating the PSA measurement at 3 and 12 months. An increment of ≥1.0 ng/mL above the corrected baseline PSA (for low testosterone and 5-alpha reductase inhibitor usage) value will lead to referral for urologic evaluation for consideration of prostate biopsy, confirmed by a repeat determination. Treatment will be discontinued for any subject who is diagnosed as having prostate cancer during the trial.

8.6.2.3. Benign prostatic hyperplasia.

Men who have evidence of moderately severe lower urinary tract symptoms, i.e., a score of >19 on the International Prostate Symptom Score (IPSS) questionnaire, will be excluded. An increase of >5 points or to an absolute value of >19 will result in a review of medications that affect urine flow rates and evaluation for prostatitis. If a cause is found, it should be treated. If no cause is found, treatment with an alpha blocker should be considered. If the subject is treated and symptoms persist, or if acute urinary retention occurs, the subject should be referred for urological consultation. If the urologist treats the subject and the score does not decrease below the above threshholds, gel treatment will be discontinued

8.6.2.4. Cardiovascular disease

Men will be monitored for the occurrence of cardiovascular events during the entire course of the two-year trial. Treatment will be discontinued in men who have a myocardial infarction or stroke. The number of subjects whose treatment is discontinued for serious adverse events will be monitored and assessed with the DSMB.

8.6.2.5. Sleep apnea.

We shall exclude men who have diagnosed sleep apnea that is not being treated, and during treatment we will question men for newly diagnosed but untreated sleep apnea.

8.6.2.6. Physical and cognitive function testing.

The small risk of physical and cognitive function testing will be minimized by training the research assistants who perform the tests how to instruct the subjects how to perform the tests

properly. For the 6-minute walk, there will be a standardized protocol for warm-up and careful supervision of the subjects during the testing. The risks associated with cognitive testing are small and primarily consist of anxiety related to concerns about performance. Testers will be trained to encourage and reassure subjects that the tests are designed to be difficult for most people.

8.6.2.7. Sexual function and vitality testing.

We shall employ several means to minimize the burden of time, the minor distress of answering personal questions, and the perceived loss of confidentiality. Use of interactive voice response (IVR) for all vitality questionnaires and for the Harbor-UCLA 7-Day Questionnaire (the primary end point for sexual function) will allow the subjects to answer these questionnaires from their homes at their convenience and thereby reduce the time they spend at the trial sites. Their answers will also be anonymous this way, and not seen by trial site personnel. The subjects may refuse to answer a question that causes them discomfort or anxiety.

8.6.2.8. Time burden.

Subjects who qualify for more than one protocol will be offered the chance to participate in those for which they qualify, but they will also be told of the approximate time burden. Study staff will be taught to be mindful of a subject's fatigue and to offer to a subject who appears fatigued the chance to resume testing on another day.

8.6.2.9. Prostate biopsy.

The standard precaution that minimizes the chance of bleeding during and after a prostate biopsy is avoiding agents that impair clotting, such as aspirin, nonsteroidal anti-inflammatory agents, and herbal supplements. The standard precaution that minimizes the risk of infection is administration of antibiotics.

8.6.2.10. Risk of using excessive testosterone gel.

No risk is expected if a subject takes a greater dose of AndroGel than prescribed, and any elevation of the serum testosterone from a single larger dose would be transient, i.e., 1-2 days. If a subject takes a larger dose than prescribed chronically, it would be detected in the serum testosterone measurements at 1, 2, 3, 6, 9 and 12 months, and the dose would be lowered.

8.6.3. Clinical Management of Participants

The T Trial Investigators recognize the obligation and importance of reporting information acquired during the research study visit to the health care provider (HCP) of participants. Participants and their HCP will be notified as soon as possible if potentially serious medical problems are identified during any of the T Trial procedures, or reported during a T Trial study visit.

8.6.3.1. Notification to Health Care Provider

Participants will be asked about several specific medical and cardiovascular events at each T Trial follow-up visit. They will be asked in the appropriate lay terms if they have experienced any of the signs and symptoms of angina and transient ischemic attack. If possible, this information will be evaluated by a T Trial physician who will determine the appropriate disposition. If the participant reports that he has not informed his HCP, the T Trial staff will notify the participant's HCP (with the participant's permission) by fax or email, as soon as possible. If determined necessary, the participant will be transported to the emergency department or escorted to an urgent care hospital visit for further evaluation and/or treatment.

If the participant is unable to identify a primary care physician or HCP, the site staff will identify one within the site's medical institution.

8.6.4. Data and Safety Monitoring Board.

An external DSMB will be established to monitor all aspects of the study. The Board will consist of experts in geriatrics, biostatistics, clinical trials, endocrinology, and prostate disease. The DSMB members will not be affiliated with the study and will be appointed by the NIA Director in consultation with the principal investigator. The Board will meet every six months to review subjects' safety, study progress and data integrity and completeness. After each meeting, the DSMB will provide the NIA Director with its recommendations, and the Director will decide whether or not to accept them.

9. Data Management

9.1. Data Management System

The Data Coordinating Center (DCC) at the University of Pennsylvania will develop a data management system for the collection, storage and management of data. This system will be developed using Oracle Corporation's suite of pharmaceutical applications. The data management systems will use a combination of tools to perform the following study functions:

- Subject tracking – to monitor recruitment and provide visit schedules for subjects and composite visit schedules for clinical sites.
- Eligibility determination - to evaluate screening data (serum testosterone, PSA, etc.) to determine eligibility for one or more efficacy areas.
- Treatment allocation - to allocate subjects to receive testosterone or placebo and to balance the treatment groups based on the minimization technique.
- Dose modification – to identify out-of-range testosterone levels.
- Specimen tracking- to document specimens from collection and processing to storage and retrieval.

9.2. Data Entry

Electronic data entry will be used primarily to achieve accuracy and efficiency. The following methods will be utilized:

- Remote data capture will permit authorized personnel to enter data remotely via a secure Internet connection.
- Electronic data transfer methods will be developed and tested to ensure that data are completely and accurately transmitted. This will include data transferred from the central laboratory and associated reading centers, as well as data collected via the Interactive Voice Response System (IVRS).

9.3. Data Quality

Oracle Clinical includes a data quality module to identify incorrect data based on a set of rules that describe the expected data. The DCC will collaborate with the investigative team to establish these parameters for primary and secondary outcomes, safety, regulatory, and descriptive values. The data management team will develop a data validation plan, rule set specifications, and programming logic to implement data validation rules. The DCC staff will interact with clinical site staff to verify queried data and track all queries to resolution.

9.3.1. Quality Control Activities

The Quality Control Committee and the DCC will develop a quality assurance and control plan that ensures that study data are as precise and reliable as possible.

Manual of Procedures (MOP) - The MOP will describe the sequence of study conduct and provide detailed instruction for the performance of screening, baseline, enrollment, treatment allocation and follow-up procedures. The MOP will provide instruction in case report form completion, use of the electronic data management system, and collection, documentation and transfer of specimens and tests to laboratories and reading centers.

Training and certification procedures - The DCC will conduct a training session before the study starts to train and certify personnel in the performance of study procedures.

Site visits – Findings from site visits will be used to resolve problems and develop corrective action plans.

External data sources - The DCC will monitor quality control of data received from study laboratories and reading centers.

Internal quality control procedures - A data validation plan, rule set specifications, and programming logic to implement data validation rules will be implemented.

9.3.2. Routine reports

The DCC will develop a set of standard enrollment, tracking, quality review, and safety monitoring reports. Adverse event reports, DSMB reports and reports for statistical analysis will be developed and produced on an appropriate schedule.

9.4. Data Security

The data management system will be designed to prevent unauthorized access to trial data and to prevent data loss due to equipment failure or catastrophic events. The procedures to do so encompass user account management, user privilege assignment, data loss prevention (database backup), computer systems validation, performance monitoring, and DMS change management. User access will be controlled by assignment of confidential usernames, passwords and role assignment. The system will meet the applicable Federal regulatory requirements and those described in the E6 Good Clinical Practice Guidelines to ensure the confidentiality of trial subjects.

Study data collected at the clinical sites will be entered into a web based data management system. This data management system uses a secure connection between the client browser at the clinical site and the web server at the DCC. Data transmitted over this connection is authenticated by the use of digital certificates and is encrypted as it travels the Internet to the DCC.

Electronic files containing data from hand held devices, the central laboratory, or the central reading center will be transferred to the DCC using secure FTP technology. The DCC team will maintain a secure FTP server. The files transmitted using this method will be encrypted during the exchange.

The DCC project team will collaborate with the Investigational Drug Service (IDS) and the biostatistics team to protect the blinding of treatment assignments and electronic access to information that could indirectly or directly lead to unblinding treatment assignment or codes. Internal access to such information is stored in password-protected files. Documentation is

stored in the locked files of the IDS at the University of Pennsylvania. Within the DCC this information is locked in files to which only department managers have access.

9.4.1. Maintaining Anonymity of Submitted Medical Records

Clinical site personnel will de-identify all medical records before sending them to the DCC by obliterating any Protected Health Information (PHI). Upon receipt, DCC personnel will review the records to ensure that no PHI is visible.

9.4.2. Confidentiality

Information about study subjects will be kept confidential and managed according to the requirements of the Health Insurance Portability and Accountability Act (HIPAA) of 1996.

Subjects will be asked to provide their Social Security Number (SSN) for the purpose of tracking their status in the National Death Index in the event they become lost to follow-up. This information will be locked in a secure location with access limited to the TTrial staff only. It will not be entered or stored in the electronic system and will be used for this purpose only. Subjects may refuse to provide this information without consequence to their study participation.

References

1. Vermeulen A, Rubens R, Verdonck L 1972 Testosterone secretion and metabolism in male senescence. J Clin Endocrinol Metab 34:730-735
2. Feldman HA, Longcope C, Derby CA, Johannes CB, Araujo AB, Coviello AD, Bremner WJ, McKinlay JB 2002 Age trends in the level of serum testosterone and other hormones in middle-aged men: longitudinal results from the Massachusetts male aging study. J Clin Endocrinol Metab 87:589-598
3. Harman SM, Metter EJ, Tobin JD, Pearson J, Blackman MR 2001 Longitudinal effects of aging on serum total and free testosterone levels in healthy men. Baltimore Longitudinal Study of Aging. J Clin Endocrinol Metab 86:724-731
4. Morley JE, Kaiser FE, Perry HM, 3rd, Patrick P, Morley PM, Stauber PM, Vellas B, Baumgartner RN, Garry PJ 1997 Longitudinal changes in testosterone, luteinizing hormone, and follicle-stimulating hormone in healthy older men. Metabolism 46:410-413
5. Jette AM, Branch LG 1981 The Framingham Disability Study: II. Physical disability among the aging. Am J Public Health 71:1211-1216
6. Fiatarone MA, Marks EC, Ryan ND, Meredith CN, Lipsitz LA, Evans WJ 1990 High-intensity strength training in nonagenarians. Effects on skeletal muscle. JAMA 263:3029-3034
7. Wolfson L, Whipple R, Derby C, Judge J, King M, Amerman P, Schmidt J, Smyers D 1996 Balance and strength training in older adults: intervention gains and Tai Chi maintenance. J Am Geriatr Soc 44:498-506
8. Morley JE 2003 Mobility performance: a high-tech test for geriatricians. J Gerontol A Biol Sci Med Sci 58:712-714
9. O'Donnell AB, Travison TG, Harris SS, Tenover JL, McKinlay JB 2006 Testosterone, dehydroepiandrosterone, and physical performance in older men: results from the Massachusetts Male Aging Study. J Clin Endocrinol Metab 91:425-431
10. Schaap LA, Pluijm SM, Smit JH, van Schoor NM, Visser M, Gooren LJ, Lips P 2005 The association of sex hormone levels with poor mobility, low muscle strength and incidence of falls among older men and women. Clin Endocrinol (Oxf) 63:152-160
11. Bhasin S, Storer TW, Berman N, Yarasheski KE, Clevenger B, Phillips J, Lee WP, Bunnell TJ, Casaburi R 1997 Testosterone replacement increases fat-free mass and muscle size in hypogonadal men. J Clin Endocrinol Metab 82:407-413
12. Snyder PJ, Peachey H, Berlin JA, Hannoush P, Haddad G, Dlewati A, Santanna J, Loh L, Lenrow DA, Holmes JH, Kapoor SC, Atkinson LE, Strom BL 2000 Effects of testosterone replacement in hypogonadal men. J Clin Endocrinol Metab 85:2670-2677
13. Kenny AM, Prestwood KM, Gruman CA, Marcello KM, Raisz LG 2001 Effects of transdermal testosterone on bone and muscle in older men with low bioavailable testosterone levels. J Gerontol A Biol Sci Med Sci 56:M266-272
14. Page ST, Amory JK, Bowman FD, Anawalt BD, Matsumoto AM, Bremner WJ, Tenover JL 2005 Exogenous testosterone (T) alone or with finasteride increases physical performance, grip strength, and lean body mass in older men with low serum T. J Clin Endocrinol Metab 90:1502-1510
15. Snyder PJ, Peachey H, Hannoush P, Berlin JA, Loh L, Lenrow DA, Holmes JH, Dlewati A, Santanna J, Rosen CJ, Strom BL 1999 Effect of testosterone treatment on body composition and muscle strength in men over 65 years of age. J Clin Endocrinol Metab 84:2647-2653
16. Bacon CG, Mittleman MA, Kawachi I, Giovannucci E, Glasser DB, Rimm EB 2003 Sexual function in men older than 50 years of age: results from the health professionals follow-up study. Ann Intern Med 139:161-168

17. Feldman HA, Goldstein I, Hatzichristou DG, Krane RJ, McKinlay JB 1994 Impotence and its medical and psychosocial correlates: results of the Massachusetts Male Aging Study. J Urol 151:54-61
18. Laumann EO, Paik A, Rosen RC 1999 Sexual dysfunction in the United States: prevalence and predictors. JAMA 281:537-544
19. Isidori AM, Giannetta E, Gianfrilli D, Greco EA, Bonifacio V, Aversa A, Isidori A, Fabbri A, Lenzi A 2005 Effects of testosterone on sexual function in men: results of a meta-analysis. Clin Endocrinol (Oxf) 63:381-394
20. Barrett-Connor E, Von Muhlen DG, Kritz-Silverstein D 1999 Bioavailable testosterone and depressed mood in older men: the Rancho Bernardo Study. J Clin Endocrinol Metab 84:573-577
21. Seidman SN, Araujo AB, Roose SP, Devanand DP, Xie S, Cooper TB, McKinlay JB 2002 Low testosterone levels in elderly men with dysthymic disorder. Am J Psychiatry 159:456-459
22. Orengo CA, Fullerton L, Kunik ME 2005 Safety and efficacy of testosterone gel 1% augmentation in depressed men with partial response to antidepressant therapy. J Geriatr Psychiatry Neurol 18:20-24
23. Seidman SN, Spatz E, Rizzo C, Roose SP 2001 Testosterone replacement therapy for hypogonadal men with major depressive disorder: a randomized, placebo-controlled clinical trial. J Clin Psychiatry 62:406-412
24. Wang C, Alexander G, Berman N, Salehian B, Davidson T, McDonald V, Steiner B, Hull L, Callegari C, Swerdloff RS 1996 Testosterone replacement therapy improves mood in hypogonadal men--a clinical research center study. J Clin Endocrinol Metab 81:3578-3583
25. Drane DL, Yuspeh RL, Huthwaite JS, Klingler LK 2002 Demographic characteristics and normative observations for derived-trail making test indices. Neuropsychiatry Neuropsychol Behav Neurol 15:39-43
26. Lamar M, Resnick SM, Zonderman AB 2003 Longitudinal changes in verbal memory in older adults: distinguishing the effects of age from repeat testing. Neurology 60:82-86
27. Schaie KW, Willis SL 1993 Age difference patterns of psychometric intelligence in adulthood: generalizability within and across ability domains. Psychol Aging 8:44-55
28. Almeida OP, Waterreus A, Spry N, Flicker L, Martins RN 2004 One year follow-up study of the association between chemical castration, sex hormones, beta-amyloid, memory and depression in men. Psychoneuroendocrinology 29:1071-1081
29. Green HJ, Pakenham KI, Headley BC, Yaxley J, Nicol DL, Mactaggart PN, Swanson C, Watson RB, Gardiner RA 2002 Altered cognitive function in men treated for prostate cancer with luteinizing hormone-releasing hormone analogues and cyproterone acetate: a randomized controlled trial. BJU Int 90:427-432
30. Salminen EK, Portin RI, Koskinen AI, Helenius HY, Nurmi MJ 2005 Estradiol and cognition during androgen deprivation in men with prostate carcinoma. Cancer 103:1381-1387
31. Ferrucci L, Maggio M, Bandinelli S, Basaria S, Lauretani F, Ble A, Valenti G, Ershler WB, Guralnik JM, Longo DL 2006 Low testosterone levels and the risk of anemia in older men and women. Arch Intern Med 166:1380-1388
32. Denny SD, Kuchibhatla MN, Cohen HJ 2006 Impact of anemia on mortality, cognition, and function in community-dwelling elderly. Am J Med 119:327-334
33. Culleton BF, Manns BJ, Zhang J, Tonelli M, Klarenbach S, Hemmelgarn BR 2006 Impact of anemia on hospitalization and mortality in older adults. Blood 107:3841-3846
34. Crawford ED, Eisenberger MA, McLeod DG, Spaulding JT, Benson R, Dorr FA, Blumenstein BA, Davis MA, Goodman PJ 1989 A controlled trial of leuprolide with and without flutamide in prostatic carcinoma. N Engl J Med 321:419-424

35. Fowler JE, Jr., Whitmore WF, Jr. 1981 The response of metastatic adenocarcinoma of the prostate to exogenous testosterone. J Urol 126:372-375
36. Thompson IM, Pauler DK, Goodman PJ, Tangen CM, Lucia MS, Parnes HL, Minasian LM, Ford LG, Lippman SM, Crawford ED, Crowley JJ, Coltman CA, Jr. 2004 Prevalence of prostate cancer among men with a prostate-specific antigen level < or =4.0 ng per milliliter. N Engl J Med 350:2239-2246
37. Snyder PJ, Peachey H, Hannoush P, Berlin JA, Loh L, Holmes JH, Dlewati A, Staley J, Santanna J, Kapoor SC, Attie MF, Haddad JG, Jr., Strom BL 1999 Effect of testosterone treatment on bone mineral density in men over 65 years of age. J Clin Endocrinol Metab 84:1966-1972
38. Thompson IM, Ankerst DP, Chi C, Lucia MS, Goodman PJ, Crowley JJ, Parnes HL, Coltman CA, Jr. 2005 Operating characteristics of prostate-specific antigen in men with an initial PSA level of 3.0 ng/ml or lower. JAMA 294:66-70
39. Parekh DJ, Ankerst DP, Higgins BA, Hernandez J, Canby-Hagino E, Brand T, Troyer DA, Leach RJ, Thompson IM 2006 External validation of the Prostate Cancer Prevention Trial risk calculator in a screened population. Urology 68:1152-1155
40. Smith G, Ivnik RJ, Petersen RC, Malec JF, Kokmen E, Tangalos E 1991 Age-associated memory impairment diagnoses: problems of reliability and concerns for terminology. Psychol Aging 6:551-558
41. Crawford JR, Henry JD 2004 The positive and negative affect schedule (PANAS): construct validity, measurement properties and normative data in a large non-clinical sample. Br J Clin Psychol 43:245-265
42. Steidle C, Schwartz S, Jacoby K, Sebree T, Smith T, Bachand R 2003 AA2500 testosterone gel normalizes androgen levels in aging males with improvements in body composition and sexual function. J Clin Endocrinol Metab 88:2673-2681
43. Cella D, Eton DT, Lai JS, Peterman AH, Merkel DE 2002 Combining anchor and distribution-based methods to derive minimal clinically important differences on the Functional Assessment of Cancer Therapy (FACT) anemia and fatigue scales. J Pain Symptom Manage 24:547-561
44. Cella D, Yount S, Sorensen M, Chartash E, Sengupta N, Grober J 2005 Validation of the Functional Assessment of Chronic Illness Therapy Fatigue Scale relative to other instrumentation in patients with rheumatoid arthritis. J Rheumatol 32:811-819

Appendix A – Questionnaire and Procedure Schedule

T Trial Questionnaires & Procedures [Revised 8/27/10]				Treatment Phase (M = Month)								Follow-up	
Tests or Procedures in All Subjects:	Screen 2	† After SV2	Base-line	M 1 & M 2	M 3	M 4 & M 5	M 6	M 7 & M 8	M 9	M 10 & M 11	M12	M18	M24
(Visit Type: V = Visit to clinic, P = phone, ** by IVR)	V		V	V	V	P	V	P	V	P	V	P/V	P
Informed consent			X										
Eligibility determination (Post-SV2) & Randomization (Baseline)	X		X										
Blood draw	X		X	X	X		X		X		X	X	
- Testosterone level	X		X	X	X		X		X		X		
- PSA			X		X						X	X	
- Hemoglobin/hematocrit	X		X		X		X		X		X		
Urinalysis (SV2 only) & stored urine (Baseline – M12)	X		X	X	X		X		X		X		
Concomitant Medication	X		X	X	X	X	X	X	X	X	X		
Medical history	X												
Adverse events/recent medical events				X	X	X	X	X	X	X	X	X	X
Cardiovascular symptoms and events			X	X	X	X	X	X	X	X	X	X	X
Gel instruction & review			X	X	X	X	X	X	X	X			
Medication Compliance				X	X		X		X		X		
Weight, waist, hip, blood pressure measures, [Ht.(SV 2 and M12 only)]	X		X		X		X		X		X		
Rectal exam (DRE)	X				X						X		
International Prostate Symptom Score (IPSS)	X		X		X						X		
All Subjects – Primary End Points													
6 Minute Walk Test	X		X		X		X		X		X		
UCLA Sex Diary – Question 4** (+ entire diary after SV2 only)		X+			X		X		X		X		
FACIT Fatigue**	X				X		X		X		X		
WMS-R-LM-II (verbal memory - paragraph recall)			X				X				X		
PANAS*		X			X		X		X		X		
Falls			X		X		X		X		X		
PHQ-9** (on paper at SV2 followed by IVR)	X	X*			X		X		X		X		
DISF-M-II	X												
3MSE			X								X		
MMSE	X												
MAC-Q			X				X				X		
SF-36			X										
Trial specific - Global Impression questions**		X			X		X		X		X		
General - Global Impression question**		X			X		X		X		X		
Tests in Subjects in Specific Trials: Secondary End Points (except PF-10-all men)													
Physical: PF-10			X		X		X		X		X		
Sexual: UCLA Sex Diary** (complete), IIEF, DISF-M-II			X		X		X		X		X		
Vitality: SF-36-Vitality**			X		X		X		X		X		
Cognitive: BVRT, Card rotation, TMT			X				X				X		

T Trials - Summary of Protocol Amendments

Amendment	Date of Amendment/ Protocol version	Summary of Changes	Rationale for Change
Amendment 20	July 16, 2013 Protocol v7.3 No consent changes	Added a follow-up lab draw for men who had elevated hemoglobin at M12 Visit. Lab draw to be completed 3 months after M12. If the M15 value is still high, we recommend referral for other causes of erythrocytosis	**Rationale:** Previously there was no specific follow-up procedure in the protocol for men with elevated hemoglobin at M12.
Amendment 19	March 15, 2013 Protocol v7.2 No consent changes	Removed 2 ml blood draw at month 12 for HgbA1c. Consent did not need to be updated since it was amended to specify HgbA1c will be tested at M12 and this is not incorrect.	**Rationale:** Removed the additional 2 ml drawn at month 12 which was added in amendment 18 because this was always being drawn. Already have HgbA1c at month 12.
Amendment 18	January 23, 2013 Protocol v7.1 Consent v7.1	Added a 2 mL blood draw to the month 12 visit for HgbA1c. Allowed enrollment to re-start/resume in sexual function trial if man is eligible for bone, cv or some combination of the two. New prescribing info for AndroGel submitted and risks were updated in the consent.	**Rationale:** We want to be able to compare what effect testosterone may have on HgbA1c and already have a pre-treatment measure. In order to complete enrollment in the bone and cv trials in a timely manner we are allowing men to enroll in Bone and/or CV if they are eligible for sexual function. The new risks in the prescribing info were added to the consent form to further inform participants.
Amendment 17	No change to protocol or consent- added 2 new epigenetics consents	Added a screening epigenetics consent and a consent at month 6 (after 6 months of T use) to enroll men into an epigenetics study - optional participation	**Rationale:** Decided by SC that it would be helpful to focus on targeted genes and have a control population (men excluded due to T level being too high) to compare to those men enrolled in the T Trial
Amendment 15/16	October 29, 2012 Protocol v 6.0 and Protocol v 7.0	Updated screening and baseline consents, as well as protocol, to move around screening and baseline procedures so men can be assessed for eligibility in CV and Bone trials sooner in the trial, since this amendment makes it mandatory that men be eligible for one of the two in order to be enrolled in the T Trial. Previously these studies were optional.	**Rationale:** In order to complete recruitment in all our ancillary studies, with the closer of the main trial 3 sub trials, we have made it now mandatory for a man to be eligible for bone or CV in order to be enrolled in the trial. This will allow us to complete enrollment in the bone and cv trials, while over enrolling as little as possible.
Amendment 14	August 13, 2012 Protocol v. 5.0	Permission received to continued enrollment in the physical and vitality trials beyond target to continue enrollment in the bone and CV trials.	**Rationale:** In order to finish enrolling in both ancillary studies it was necessary to overall, beyond the expected targets, for the main trials (sexual, physical and vitality). This is similar to amendment 11.
Amendment 13	June 6, 2012 Protocol v4.0	Increase the max. allowable dose of AndroGel from 10 grams to 15 grams.	**Rationale:** Review of T Trial data led to realization that many men in the trial, with T levels below 500ng/dL, were already at the max. allowable dose of 10 grams per day. The T Trial team petitioned the FDA to increase the daily dose to 15 grams in order to take further measures to increase testosterone levels among participants. This was approved by the FDA as long as participant who are increased to 15 grams have a follow up blood draw.

T Trials - Summary of Protocol Amendments

Amendment	Date of Amendment/ Protocol version	Summary of Changes	Rationale for Change
Amendment 12	March 14, 2012 Protocol v3.8	Increase the limit of testosterone at which we administer dose increases from 400ng/dL to 500ng/dL.	**Rationale:** A review of aggregated serum testosterone concentrations in subjects using testosterone gel showed that about one third of them have values less than 400 ng/dL, the lower limit of the target range, at each visit during the treatment phase of the trial. In addition, the median testosterone concentration during treatment is only about 400 ng/dL, rather than in the middle of the target range of 400-800 ng/dL. These values indicate that the Trial is not achieving its goal of increasing the serum testosterone concentration of elderly men to that of young men. The Steering Committee therefore decided to increase the lower limit of the target range from 400 to 500 ng/dL.
Amendment 11	February 17, 2012 Protocol v3.7 (baseline consent version 6.0 and prostate consent)	Specified that once a trial reaches its target enrollment will continue to enroll men in that trial only if they are eligible for one of the other trials, which has not reached target enrollment. Submitted prostate tissue additional samples consent form and manual. Updated consent form to only exclude men from trial with clinically apparent CV disease	**Rationale:** Protocol to collect additional slices when men have clinically warranted need for prostate biopsy was never submitted to site IRBs. Sent in with this amendment. Sexual function trial reached target enrollment. Continue to enroll men in it if they are eligible for other trials.
Amendment 10	December 7, 2011 Protocol v3.6	Specified need for self-reported decreased libido within protocol for inclusion into the sexual function trial Specified that testosterone or placebo use will stop if testosterone is > 800 ng/dL after 2 consecutive dose reductions.	**Rationale:** Both of these changes were made to wording in the protocol to be consistent with other sections within the protocol.
Amendment 9	*August 1, 2011 Protocol v 3.5 (no changes to the baseline consent)	Change to physical function inclusion criteria- change gait speed from <1.0m/s to 1.2 m/s.	**Rationale:** Enrollment in the physical function trial has been much slower than expected and if it continues at the current pace enrollment will not finish until late in 2017. If the gait speed criteria were to be relaxed slightly we would be able to immediately enroll 43 men within the physical function trial and the projected end date of the trial would be 6 months after the vitality trial is projected to be completed.

T Trials - Summary of Protocol Amendments

Amendment	Date of Amendment/ Protocol version	Summary of Changes	Rationale for Change
Amendment 8	June 23, 2011 Protocol v 3.4 (baseline consent v 5.0)	Under the current protocol the secondary endpoints for the physical function trial (the PF 10 and the physical function impression of change question) are only asked of men in the physical function trial. This amendment changed the protocol so that these questionnaires would be asked of all men enrolled in the trial.	**Rationale:** The steering committee determined asking these questionnaires would provide the study with usual data on physical function perception and will provide no additional burden to subjects. Both questionnaires are very short and not at all sensitive in nature.
Amendment 7	April 19, 2011 Protocol v 3.3 Baseline Consent v 4.0	Change the study sponsor from Solvay to Abbott and Modify the low grade prostate cancer risk criteria from < 30% to < 35%	**Rationale:** No prior clinical trial has used the prostate risk calculator and it is considered to be a stringent measure of risk- high grade prostate cancer risk will remain at <7% and it is believed changing the low grade risk from <30% to < 35% will not adversely affect participants
		Change the cut off level for repeating testosterone from >2000 ng/dL to >1500ng/dL	**Rationale:** At the request of the DSMB changed this from testosterone levels >200 0 to >1500
		Allow men who discontinue treatment to resume treatment (unless participant has an MI or stroke) after 2 weeks' time if their testosterone level is <400 ng/dL	**Rationale:** Allowing participants to resume gel use will help with protocol compliance and not adversely affect safety of subjects since men will only be allowed to restart at one depression and will have a testosterone level at the lowest end of the normal range <400ng/dL
		Clarified confusing IPSS criteria/text	**Rationale:** Updated monitoring plans to make them more consistent throughout the trial
		Refined hemoglobin monitoring- incorporated into dose adjustment schema and repeat Hgb before referring participant	**Rationale:** Repeat the testosterone level and Hgb, if the Hgb is still elevated on repeat test and after testosterone is reduced, refer participants. Believe this does not reduce patient safety but will cut down on unnecessary referrals.
		Add creatinine to the baseline visit and Add eGFR calculation to the SV1 visit	**Rationale:** Include creatinine and eGFR calculation for assessing eligibility of men for the CV trial

Amendment	Date of Amendment/ Protocol version	Summary of Changes	Rationale for Change
		• Consent Form Changes: • Change study sponsor from Solvay to Abbott • Include in the consent that a certificate of confidentiality has been obtained • Change the low grade prostate risk explanation from <30% to < 35% • Add sentence about not taking testosterone from another source while you are in this trial as it could harm participants and confuse interpretation of the final study results	**Rationale:** Same as above- consent form was updated to match protocol
Amendment 6	November 29,2010 Protocol v3.2 Consent v3.0	After monitoring slow enrollment in the physical function trial for several months, changes were proposed to increase the yield in the physical function trial. Following the in person Steering Committee Meeting in November 2010, some additional changes were proposed: The initial question used to asses subjective physical function complaints was: "Do you have difficulty walking a quarter of a mile? Yes or no?	Increase enrollment into the physical function trial while still using questions which are considered standard/ validated for enrollment into physical function trials
		Consultation with physical function exports led the development of 2 new questions to measure physical function difficulty: 1. During the past month, how much difficulty have you had walking ¼ of a mile (about 3-4 city blocks) without the aid of another person or any special equipment? (Special equipment is defined as use of a cane, or a walker. It does not include a motorized walker, wheelchair or any motorized device.) No difficulty A little difficulty Some difficulty A lot of difficulty Unable to do it Do not do it for reasons other than difficulty Do not know, or refused to answer	

Amendment	Date of Amendment/ Protocol version	Summary of Changes	Rationale for Change
		2. During the past month, how much difficulty have you had climbing one flight of stairs without the aid of another person or any special equipment? No difficulty A little difficulty Some difficulty A lot of difficulty Unable to do it Do not do it for reasons other than difficulty Do not know, or refused to answer	
		Subjects will qualify to proceed in the Trial if they answer "a little difficulty," "some difficulty," "a lot of difficulty," or "unable to do it" to either question. Exclude men with a serum TSH >7.5 mIU/L at Screening Visit 2 Developed a procedure for repeating testosterone labs for men who have a testosterone level >2000 ng/dL and re-educating men as to how to apply the gel as this value most likely resulted from improper gel application	
		Simplified the cancer exclusion criteria to make it easier to operationalize from the original criteria to exclude, "Breast and lung cancer, and cancers that limit life expectancy to <5 years, or which require current therapy will be excluded. However, nonmelanotic skin cancers or other cancers that are deemed cured or stable without current therapy (e.g., treated well differentiated thyroid cancers, testicular cancers, Hodgkin's lymphoma, and colon cancer) may be enrolled at the investigator's discretion. " Changed the language to: "Exclude all men who have been diagnosed or treated for cancer in the past 3 years."	Make it easier to operationalize inclusion/exclusion criteria related to cancer diagnosis and treatment while still using an exclusion criteria which is considered standard by most research oncologists

Amendment	Date of Amendment/ Protocol version	Summary of Changes	Rationale for Change
		Revised the psychiatric exclusion criteria, which was generic and made a blanket statement to "exclude all men with Axis I psych disorders" This was changed to: "exclude men who have major psychiatric disorders, including major depression (PHQ-9 score >14), mania, hypomania, psychosis, schizophrenia or schizoaffective disorders, that are untreated, unstable, have resulted in hospitalization or medication change within the previous three months, or would result in inability to complete the trial efficacy parameters. Subjects whose disorders have been stable while being treated for more than three months are eligible."	Allow enrollment of men who have been diagnosed with psychiatric disorders in the past but are considered to be stable and well treated
		Consent changes: Changes were made to specify in the consent men may be brought back for a repeat blood draw to check lab values for safety concerns, irregular values, etc. and this may occur in those on active or on placebo gel	
Amendment 5	September 13, 2010 Protocol v 3.1 (no changes to the baseline consent)	Change BMI exclusion criterion from >35 kg/m^2 to >37 kg/m^2.	**Rationale:** 8% of screened men were being excluded due to a BMI >35 kg/m^2; almost all of these BMIs were between 35 and 37 kg/m^2. Since large changes in BMI are needed to affect testosterone to a measurable degree raising the BMI upper limit slightly will not meaningfully affect serum testosterone concentration
Amendment 4	June 22, 2010 Protocol v 3.0 (no changes to the baseline consent) SVX consent developed	Modification of testosterone entry level criteria: Original entry criteria: both Screening Visit testosterone concentrations between 100 and 250 ng/dL, inclusive. Revised criteria: SV1 testosterone <275 ng/dL SV2 testosterone <300 ng/dL Mean of SV1 and SV2 testosterone <275 ng/dL Men with testosterone <100 with other labs in any of the following ranges will be referred for further testing by MRI. serum LH > 9.3 mIU/mL total T4 < 4.5 µg/dL prolactin >30 ng/mL cortisol <10 µg/dL)	**Rationale:** enlarge yield from screening while still maintaining a strict upper cap on testosterone level and ensuring subject safety

Amendment	Date of Amendment/ Protocol version	Summary of Changes	Rationale for Change
Amendment 3	April 8, 2010 Protocol v 2.0 Baseline Consent v 3.0	Addition of two exclusion criteria: 1) Stroke within the previous three months and 2) systolic blood pressure of >160 mm Hg or a diastolic blood pressure >100 mm Hg.	
		Reduce **starting dose** of AndroGel from 7.5 to 5.0 grams a day	**Rationale:** excluding men at higher risk for heart disease.
		Reduce **target on-treatment range** of serum testosterone from 500-900 ng/dL to 400-800 ng/dL.	**Rationale:** reduce chance of higher than normal serum testosterone concentration during the first month of treatment.
		Discontinue treatment following myocardial infarction or a stroke.	**Rationale:** 400-800 ng/dL is the normal physiologic range in young men.
		Administer questionnaires about **cardiovascular health** at baseline and during treatment.	**Rationale:** these men are at higher risk of another cardiovascular event.
		Lower the hemoglobin A1c exclusion criteria from >9% to >8.5%	**Rationale:** minimize risk of cardiovascular events; obtain information on possible association of testosterone with such events
		Standardize blood draws at screening: Blood must be drawn between 7 – 10 AM. Blood should be collected from subjects who have been fasting, defined as drinking only water in the previous 8 hours. Only fasting samples are acceptable.	**Rationale:** reduce variability of T measurements.
		Consent Form Changes: Add possibility that taking testosterone may increase risk of cardiovascular events. Request permission to contact health care providers about important health information.	

VI. Cognitive Function Trial

VI.A Primary Endpoint

VI.A.1 Primary Analysis. The primary analysis will consist of an intent-to-treat comparison of the change in responses to the Wechsler Memory Scale Revised (WMS-R) Logical Memory II from baseline in randomized subjects across all trials demonstrating symptoms of Age-associated Memory Impairment (AAMI) at baseline, where AAMI is defined by a score that is more than one SD below the performance for men aged 20-24 on the WMS-R or Benton Visual Retention Test. The WMS-R LM II involves a delayed paragraph recall activity scored in two components, each ranging from 0-25. The final score is the sum of each component, therefore falling in the range 0-50. WMS-R LM II scores will be treated as continuous with change compared between treatment arms using linear random effects models adjusting for study site, indicator variables of participation in each primary efficacy trial, baseline testosterone concentration (<200), age (≤ 75), use of anti-depressants, use of PDE-inhibitors, baseline WMSR, categorical education , and version of the WMSR. Time will be modeled categorically, with an indicator for month 12 included as a covariate and an interaction of treatment and time to allow for a non-constant treatment effect. If the interaction term is not significant at the 0.05 level, results from the main effect model without the interaction will be included. Subjects with a baseline and at least one post-baseline WMS-R assessment will be included in the primary analysis. The two-sided Wald test p-value and confidence interval will be reported.

VI.A.2 Sensitivity Analysis for Missing Data. The primary analysis yields valid inference under the missing-at-random (MAR) assumption. To assess the sensitivity of results to non-ignorable missingness, shared parameter models and pattern-mixture models will be used. Details are discussed in section II.A.2.

VI.B Secondary Endpoints. All secondary endpoints will be analyzed only on subjects demonstrating AAMI and adjusting for education and balancing factors.

VI.B.1.1 Benton Visual Retention Test (BVRT). The BVRT will be scored according to the number of 'Figures with Errors' following the approach taken in the Women's Health Initiative Study of Cognitive Aging. Under this approach, the maximum possible score may exceed 26. Change in BVRT scores from baseline will be treated as continuous and compared between AAMI Androgel and placebo subjects using linear random effects models adjusting for balancing factors as described in the primary analysis. To improve interpretability in observed outcomes, BVRT scores will be negated prior to transformation and defining change in defining the regression model. As with the primary endpoint analysis, an interaction with time will be included in the model, and if not significant, results from the main effect model will be presented.

VI.B.1.2 Card Rotations Test. Change in performance on the Card Rotations Test will be analyzed using linear random effects models adjusting for baseline performance, balancing factors,

education, and test version. The test consists of a series of 10 primary figures, each of which has 8 corresponding secondary figures. Subjects are asked to determine which of the secondary figures is the same as the corresponding primary figure, and the score is taken as the number of figures answered correctly minus the number of figures answered incorrectly. The maximum score is 80 for subjects who answer all items correctly.

VI.B.1.3 Trail Making Test. The Trail Making test will be analyzed using the same methods as described in VI.B.1.2. In this assessment, subjects are asked to connect a set of numbers (Part A) or letters and numbers (Part B) in sequential order. The score for each part is the total time (in seconds) to complete both parts. The outcome analyzed will be the log transformed total time for Trails A minus the log-transformed total time for Trails B to provide a measure of working memory. Under this construction more positive scores indicate better performance.

VI.B.1.4 3MSE. Scores on the Modified Mini-Mental State Examination (3MSE) range from 0-100 and will be analyzed as continuous variables. Change from baseline in the performance of AAMI Androgel versus placebo patients between baseline and month 12 will be compared using a multivariable linear model adjusting for baseline 3MSE performance, balancing factors, and education. Note that unlike the other endpoints, which are measured at 0, 6, and 12 months, the 3MSE is only measured at baseline and 12 months.

VI.B.2. Sensitivity Analysis for Missing Data. Of the Benton, Card Rotations, and 3MSE, sensitivity analyses for endpoints that are at least marginally significant (p=0.2) will be conducted. Details of these methods may be found in section II.A.2.

VI.C Exploratory Endpoints

VI.C.1.1 WMS-R.

VI.C.1.2. BVRT.

VI.C.1.3. Card Rotations.

VI.C.1.4 Trail Making Test.

VI.C.1.5 3MSE.

VI.C.1.1-VI.C.1.5 Performance on the primary and secondary endpoints will be compared between Androgel and placebo arms across all (AAMI and non-AAMI) randomized subjects. The analysis will use the same methods as the primary and secondary analyses, adjusting for AAMI status. Evaluating changes from baseline in testosterone vs. placebo arms will first consider the continuous outcomes/change scores. In the second set of analyses, all secondary outcomes will be defined as a binary variable 'improve or not' based on the number of standard deviations of change a subject has moved, and analysis will employ logistic mixed models.

VI.C.1.6 PGIC. The impact of testosterone treatment on cognitive function will also be explored by comparing responses to the Patient Global Impression of Change in memory in AAMI subjects. To facilitate modeling efforts, responses categories will be collapsed into 5 levels, where (-2,-3) and (2,3) indicate a medium/large change, -1 and 1 indicate a small change, and 0 indicates no change. Random effects proportional odds models will evaluate differences in self-reported memory for subjects randomized to Androgel vs. placebo. If proportional odds do not hold, the analogous adjacent-categories logit model will be used. Models will adjust for balancing factors and baseline measurements of response variables. A second analysis will consider all T-Trial participants using the same model.

VI.C.1.7 Effect of testosterone on transition from non-AAMI to AAMI. A tertiary variable classifying subjects as stay the same, improve, or decrease at month 12 compared to baseline will be compared between treatment arms using a proportional odds model. Change will be defined according to the standards used in the WHI. Improvement out of AAMI and worsening into AAMI will also be analyzed as dichotomous outcomes adjusting for balancing factors among patients with AAMI and patients without AAMI, respectively.

VI.C.2. Multiplicity. The analysis of the WMS-R across all (AAMI and non-AAMI) subjects in VI.C.1 will be adjusted for multiple comparisons considering the 3 other primary endpoints across all T-trial subjects. The method of Hochberg 1988 will be used to control the Family Wise Error Rate.

VI.D Subgroup Analyses. Subgroups defined by baseline variables will be evaluated by adding interactions with treatment to the primary outcome model. Interactions with the following variables listed in VI.D.1 be considered-

VI.D.1.1 Baseline T (free and total) (continuous or above and/or <200.)

VI.D.1.2. Indicator for AAMI vs. non-AAMI.

VI.D.1.3 Baseline 3MSE group, defined as a tertiary variable according to WHI definitions (Espeland et al. 2004). 3MSE categories are defined by 95-100, 89-94, and 88 or less.

VI.D.2 Achieving the target range. Methods for time-dependent covariates will evaluate differential treatment effects on those who achieved vs. did not achieve the target level as described in section II.D.2.a.