Contents:

Roles of Investigators and Others in the Testosterone Trials
Clinical Trial Sites
Data and Safety Monitoring Board Members
TTRIALS Protocol
Bone Trial Protocol
Bone Trial Analytic Plan
# ROLES OF INVESTIGATORS AND OTHERS IN THE TESTOSTERONE TRIALS

## Steering Committee

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

## Statisticians (Department of Biostatistics and Epidemiology, University of Pennsylvania)

<table>
<thead>
<tr>
<th>Name</th>
<th>Affiliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Susan S Ellenberg, PhD</td>
<td></td>
</tr>
<tr>
<td>Xiaoling Hou, MS</td>
<td></td>
</tr>
<tr>
<td>J Richard Landis, PhD</td>
<td></td>
</tr>
<tr>
<td>Liyi Cen, MD</td>
<td></td>
</tr>
<tr>
<td>Alisa Stephens-Shields, PhD</td>
<td></td>
</tr>
<tr>
<td>Bret Zeldow, MS</td>
<td></td>
</tr>
<tr>
<td>Renee H Moore, PhD</td>
<td></td>
</tr>
</tbody>
</table>

## Data Coordinating Center (Clinical Research Computing Unit, Department of Biostatistics and Epidemiology, University of Pennsylvania)

<table>
<thead>
<tr>
<th>Name</th>
<th>Affiliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denise Cifelli, MS</td>
<td></td>
</tr>
<tr>
<td>Laura Fluharty, MPH</td>
<td></td>
</tr>
<tr>
<td>Shawn Ballard, MS</td>
<td></td>
</tr>
<tr>
<td>James Dattilo, BS</td>
<td></td>
</tr>
<tr>
<td>Sandra Smith, AS</td>
<td></td>
</tr>
<tr>
<td>Darlene Dougar, MPH</td>
<td></td>
</tr>
<tr>
<td>Laura Gallagher, MPH, CCRP</td>
<td></td>
</tr>
<tr>
<td>Tracy Chai, MS</td>
<td></td>
</tr>
<tr>
<td>Trina Brown</td>
<td></td>
</tr>
<tr>
<td>Fran Chicchi, BS</td>
<td></td>
</tr>
</tbody>
</table>

## Adjudicators of Cardiovascular and Cerebrovascular Adverse Events

<table>
<thead>
<tr>
<th>Name</th>
<th>Affiliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scott E Kasner, MD</td>
<td>University of Pennsylvania</td>
</tr>
<tr>
<td>Cora E Lewis, MD, MSPH</td>
<td>University of Alabama at Birmingham</td>
</tr>
<tr>
<td>Steven R Messe, MD</td>
<td>University of Pennsylvania</td>
</tr>
<tr>
<td>Emile R Mohler III, MD</td>
<td>University of Pennsylvania</td>
</tr>
</tbody>
</table>
### Clinical Trial Sites

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

<table>
<thead>
<tr>
<th>Member</th>
<th>Affiliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Michael O. Thorner, MD</td>
<td>Department of Medicine, University of Virginia</td>
</tr>
<tr>
<td>Eric Klein, MD</td>
<td>Cleveland Clinic</td>
</tr>
<tr>
<td>Wayne A. Meikle, MD</td>
<td>University of Utah</td>
</tr>
<tr>
<td>Wayne J. Hellstrom, MD</td>
<td>Department of Urology, Tulane University</td>
</tr>
<tr>
<td>J. Philip Miller, PhD (Chair)</td>
<td>Washington University</td>
</tr>
<tr>
<td>Richard Chappell, PhD</td>
<td>Department of Biostatistics, University of Wisconsin</td>
</tr>
<tr>
<td>George A. Kuchel, MD</td>
<td>Division of Geriatrics, University of Connecticut</td>
</tr>
<tr>
<td>Manuel D. Cerqueira, MD</td>
<td>Cleveland Clinic Nuclear Medicine</td>
</tr>
<tr>
<td>E. Magnus Ohman, MD</td>
<td>Duke University Medical Center</td>
</tr>
</tbody>
</table>
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
Amended: 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
<table>
<thead>
<tr>
<th>Amended:</th>
<th>October 17, 2012</th>
<th>Version 6.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amended:</td>
<td>October 24, 2012</td>
<td>Version 7.0</td>
</tr>
<tr>
<td>Amended:</td>
<td>January 23, 2013</td>
<td>Version 7.1</td>
</tr>
<tr>
<td>Amended:</td>
<td>March 15, 2013</td>
<td>Version 7.2</td>
</tr>
<tr>
<td>Amended:</td>
<td>July 16, 2013</td>
<td>Version 7.3</td>
</tr>
</tbody>
</table>
# TABLE OF CONTENTS

1. **INTRODUCTION** .................................................................................................................. 1
   1.1. BACKGROUND .................................................................................................................. 1
   1.2. DECREASE IN TESTOSTERONE AS MEN AGE .............................................................. 1
   1.3. CONDITIONS THAT TESTOSTERONE TREATMENT MIGHT IMPROVE .............................. 1
       1.3.1. Physical Function ....................................................................................................... 1
       1.3.2. Sexual Function ........................................................................................................ 2
       1.3.3. Vitality ....................................................................................................................... 2
       1.3.4. Cognitive Function ................................................................................................... 2
       1.3.5. Anemia ..................................................................................................................... 2
   1.4. CONDITIONS TESTOSTERONE MIGHT WORSEN .......................................................... 2
       1.4.1. Prostate Cancer ......................................................................................................... 2
       1.4.2. Benign Prostatic Hypertrophy (BPH) and Lower Urinary Tract Symptoms (LUTS) ....... 3
       1.4.3. Erythrocytosis .......................................................................................................... 3
       1.4.4. Sleep Apnea .............................................................................................................. 3
       1.4.5. Cardiovascular Disease ........................................................................................... 4
   1.5. GENETIC PROPENSITY TO RESPOND TO TESTOSTERONE .......................................... 4

2. **STUDY OBJECTIVES** ........................................................................................................... 4
   2.1. PHYSICAL FUNCTION TRIAL .......................................................................................... 4
   2.2. SEXUAL FUNCTION TRIAL ............................................................................................ 5
   2.3. VITALITY TRIAL ............................................................................................................... 5
   2.4. COGNITIVE FUNCTION TRIAL ....................................................................................... 6
   2.5. ANEMIA TRIAL ................................................................................................................ 6
   2.6. MEASUREMENTS ACROSS ALL TRIALS ...................................................................... 6

3. **STUDY DESIGN** ................................................................................................................... 7
   3.1. GENERAL DESIGN ......................................................................................................... 7
   3.2. STUDY ENDPOINTS ....................................................................................................... 7
       3.2.1. Physical Function Trial Endpoints .............................................................................. 7
       3.2.2. Sexual Function Trial Endpoints ................................................................................ 8
       3.2.3. Vitality Trial Endpoints ............................................................................................ 8
       3.2.4. Cognitive Function Trial Endpoints .......................................................................... 8
       3.2.5. Anemia Trial Endpoint ............................................................................................ 9
   3.3. SAFETY MEASUREMENTS ............................................................................................. 9
       3.3.1. Prostate Cancer ........................................................................................................ 9
       3.3.2. Lower Urinary Tract Symptoms Due to Benign Prostatic Hyperplasia ....................... 9
       3.3.3. Erythrocytosis ......................................................................................................... 9
       3.3.4. Sleep Apnea ............................................................................................................ 9
       3.3.5. Cardiovascular Disease .......................................................................................... 9
       3.3.6. Fractures ............................................................................................................... 10

4. **SUBJECT SELECTION AND WITHDRAWAL** .................................................................... 10
   4.1. NUMBER OF SUBJECTS .................................................................................................. 10
   4.2. COMMON INCLUSION CRITERIA .................................................................................... 10
   4.3. COMMON EXCLUSION CRITERIA .................................................................................. 10
       4.3.1. Evaluation of T Level < 100 .................................................................................... 12
   4.4. INCLUSION AND EXCLUSION CRITERIA FOR PHYSICAL FUNCTION TRIAL ................ 12

V 7.3 20130716
4.5. Inclusion and Exclusion Criteria for Sexual Function Trial ............................................. 13
4.6. Inclusion and Exclusion Criteria for Vitality Trial ........................................................ 13
4.7. Cognition Trial .................................................................................................................. 13
4.8. Inclusion and Exclusion Criteria for Anemia Trial .......................................................... 13
4.9. Subject Recruitment and Screening ................................................................................. 13
4.10. Early Withdrawal of Subjects .......................................................................................... 14

5. STUDY DRUG ..................................................................................................................... 15
5.1. Description ............................................................................................................................ 15
5.2. Treatment Regimen ........................................................................................................... 15
5.3. Method for Assigning Subjects to Treatment Groups ...................................................... 16
5.4. Preparation and Administration of Study Drug ............................................................... 16
5.5. Storage ............................................................................................................................... 16
5.6. Dispensing of Study Drug, and Return or Destruction of Study Drug ......................... 16
5.7. Concomitant Medications ............................................................................................... 16

6. STUDY PROCEDURES AND VISITS ............................................................................... 16
6.1. Telephone Prescreening ..................................................................................................... 16
6.1.1. Screening Visit 1 ............................................................................................................. 17
6.1.2. Screening Visit 2 ............................................................................................................. 17
6.1.3. Baseline Visit ................................................................................................................ 18
6.1.4. Months 1 and 2 Visits (± 7 days) .................................................................................. 19
6.1.5. Month 3 Visit (± 2 weeks) ............................................................................................ 20
6.1.6. Months 4 and 5 Assessments (± 7 days) ....................................................................... 21
6.1.7. Month 6 Visit (± 2 weeks) ............................................................................................ 21
6.1.8. Months 7 and 8 Assessments (± 7 days) ....................................................................... 22
6.1.9. Month 9 Visit (± 2 weeks) ............................................................................................ 22
6.1.10. Months 10 and 11 Assessments (± 7 days) ................................................................. 22
6.1.11. Month 12 Visit (± 2 weeks) ........................................................................................ 23
6.1.12. Months 18 and 24 Assessments (± 1 month) ............................................................. 24
6.2. Subject Compensation ...................................................................................................... 24
6.3. Screening, Assessment, & Monitoring Schedule .............................................................. 25

7. STATISTICAL PLAN ............................................................................................................ 26
7.1. Analytical Methods and Sample Size Estimations: Overview ........................................... 26
7.1.1. Analysis of Primary and Secondary Endpoints for the Individual Trials .................... 26
7.1.2. Sample Size Estimation ............................................................................................... 26
7.1.3. Physical Function Trial ............................................................................................... 26
7.1.4. Sexual Function Trial .................................................................................................. 27
7.1.5. Vitality/Fatigue ............................................................................................................. 27
7.1.6. Cognition Trial ............................................................................................................. 27
7.1.7. Anemia Trial ................................................................................................................ 27
7.2. Analytic Plans for Measures Across All Trials ................................................................. 27
7.2.1. Efficacy Endpoints from Individual Trials ................................................................. 27
7.2.2. Patient Global Impression of Change (PGIC) ............................................................. 28
7.3. Adverse Events ................................................................................................................ 28
7.3.1. Prostate Cancer .......................................................................................................... 28
7.4. Sample Size for the Entire Study ..................................................................................... 28
7.5. Interim Monitoring .......................................................................................................... 29

8. SAFETY AND ADVERSE EVENTS ..................................................................................... 29
8.1. DEFINITIONS .............................................................................................................................. 29
  8.1.1. Adverse Event .................................................................................................................. 30
  8.1.2. Serious Adverse Event .................................................................................................... 30
  8.1.3. Unanticipated Problem ................................................................................................. 30
  8.1.4. Adverse Event Reporting Period .................................................................................. 30
  8.1.5. Preexisting Condition ................................................................................................. 30
  8.1.6. General Physical Examination Findings .................................................................... 31
  8.1.7. Post-study Adverse Event ........................................................................................... 31
  8.1.8. Abnormal Laboratory Values ....................................................................................... 31
  8.1.9. Hospitalization, Prolonged Hospitalization or Surgery .............................................. 31
  8.2. RECORDING OF ADVERSE EVENTS .................................................................................. 31
  8.3. REPORTING OF SERIOUS ADVERSE EVENTS ................................................................. 31
    8.3.1. Study Sponsor Notification by Investigator ................................................................ 31
    8.3.2. IRB Notification by Investigator .............................................................................. 32
  8.4. UNBLINDING ......................................................................................................................... 32
  8.5. STOPPING BOUNDARIES ................................................................................................. 32
  8.6. MONITORING SUBJECT SAFETY ....................................................................................... 33
    8.6.1. Potential Risks to Subjects ........................................................................................ 33
    8.6.2. Protection Against Risk ............................................................................................ 34
    8.6.3. Clinical Management of Participants ....................................................................... 36
    8.6.4. Data and Safety Monitoring Board ........................................................................... 37

9. DATA MANAGEMENT .................................................................................................................. 37
  9.1. DATA MANAGEMENT SYSTEM .......................................................................................... 37
  9.2. DATA ENTRY ......................................................................................................................... 37
  9.3. DATA QUALITY .................................................................................................................... 37
    9.3.1. Quality Control Activities ........................................................................................ 38
    9.3.2. Routine reports .......................................................................................................... 38
  9.4. DATA SECURITY .................................................................................................................. 38
    9.4.1. Maintaining Anonymity of Submitted Medical Records ........................................ 39
    9.4.2. Confidentiality .......................................................................................................... 39
# Study Summary

<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>The Testosterone Trials</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Protocol Number</strong></td>
<td>808676</td>
</tr>
<tr>
<td><strong>Study Design</strong></td>
<td>Randomized, placebo-controlled, double-blind study of five coordinated trials</td>
</tr>
<tr>
<td><strong>Study Duration</strong></td>
<td>Six years</td>
</tr>
<tr>
<td><strong>Study Centers</strong></td>
<td>Multi-center set of trials involving 12 clinical sites geographically distributed across the United States</td>
</tr>
<tr>
<td><strong>Objectives</strong></td>
<td>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.</td>
</tr>
<tr>
<td><strong>Number of Subjects</strong></td>
<td>800</td>
</tr>
<tr>
<td><strong>Diagnosis and Main Inclusion Criteria</strong></td>
<td>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, &lt; 275 ng/dL), and symptoms and objective manifestations of mobility disability, low libido, or low vitality.</td>
</tr>
<tr>
<td><strong>Study Product, Dose, Route, Regimen</strong></td>
<td>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.</td>
</tr>
<tr>
<td><strong>Duration of administration</strong></td>
<td>AndroGel or placebo will be administered to each subject for 12 months.</td>
</tr>
<tr>
<td><strong>Reference therapy</strong></td>
<td>The effects of AndroGel on the primary and secondary end points will be compared to effects of placebo on these end points.</td>
</tr>
<tr>
<td><strong>Statistical Methodology</strong></td>
<td>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.</td>
</tr>
</tbody>
</table>
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 thresholds, 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.
Testosterone Trial Protocol

- 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/m²
- 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 testosterone 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
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 mIU/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. AbbVie will provide AndroGel in pumps, which deliver 1.25 g of gel per depression. AbbVie will 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 15 g/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 by AbbVie to 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 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:
  o 6-Minute Walk Test (Physical Function Trial)
• Other Common Endpoints:
  o Patient Global Impression Questions
  o Falls
  o 3MSE, WMS-R LM II, BVRT, Card Rotations, and Trail Making Test
  o MAC-Q
  o 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:
  o PF-10
• Subjects enrolled in the Sexual Function Trial:
  o DISF-M-II SR
  o IIEF
• Subjects enrolled in the Vitality Trial:
  o 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.

Month 3 Visit – Secondary and Exploratory Efficacy Endpoints for Each Trial

- 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)
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°C
- 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

Month 6 Visit – Secondary and Exploratory Efficacy Endpoints for Each Trial

- 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

Month 9 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)
  - 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.
Testosterone Trial Protocol

- 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

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

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>
<thead>
<tr>
<th>Trial</th>
<th>One Arm</th>
<th>Both Arms</th>
<th>+5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical function</td>
<td>174</td>
<td>348</td>
<td>366</td>
</tr>
<tr>
<td>Sexual function</td>
<td>131</td>
<td>262</td>
<td>275</td>
</tr>
<tr>
<td>Vitality</td>
<td>200</td>
<td>400</td>
<td>420</td>
</tr>
<tr>
<td>All Trials</td>
<td>505</td>
<td>1010</td>
<td>1051</td>
</tr>
</tbody>
</table>

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

### 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>
<thead>
<tr>
<th>Interim Analysis</th>
<th>Nominal Alpha for Boundary</th>
<th>Hazard Ratio on Boundary</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.01</td>
<td>4.9</td>
</tr>
</tbody>
</table>
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 discontinued.
The 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/edrnci/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 thresholds, 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

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


## Appendix A – Questionnaire and Procedure Schedule

### T Trial Questionnaires & Procedures [Revised 8/27/10]

<table>
<thead>
<tr>
<th>Tests or Procedures in All Subjects:</th>
<th>Screen 2</th>
<th>After SV2</th>
<th>Base-line</th>
<th>M 1 &amp; M 2</th>
<th>M 3</th>
<th>M 4 &amp; M 5</th>
<th>M 6</th>
<th>M 7 &amp; M 8</th>
<th>M 9</th>
<th>M 10 &amp; M 11</th>
<th>M 12</th>
<th>M 18</th>
<th>M 24</th>
</tr>
</thead>
<tbody>
<tr>
<td>Informed consent</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eligibility determination (Post-SV2 &amp; Randomization (Baseline)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blood draw</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Testosterone level</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- PSA</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Hemoglobin/hematocrit</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urinalysis (SV2 only) &amp; stored urine (Baseline – M12)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Concomitant Medication</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medical history</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adverse events/recent medical events</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cardiovascular symptoms and events</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gel instruction &amp; review</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medication Compliance</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weight, waist, hip, blood pressure measures, [Ht.(SV 2 and M12 only)]</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rectal exam (DRE)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>International Prostate Symptom Score (IPSS)</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>All Subjects – Primary End Points</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 Minute Walk Test</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UCLA Sex Diary – Question 4** (+ entire diary after SV2 only)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FACIT Fatigue**</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WMS-R-LM-II (verbal memory - paragraph recall)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PANAS*</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Falls</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHQ-9** (on paper at SV2 followed by IVR)</td>
<td>X</td>
<td>X*</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DISF-M-II</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3MSE</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MMSE</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAC-Q</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SF-36</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trial specific - Global Impression questions**</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General - Global Impression question**</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Tests in Subjects in Specific Trials: Secondary End Points (except PF-10-all men)

| Physical: PF-10                      | X        | X         | X         | X         | X   | X         |     |           |     |           |      |      |      |
| Sexual: UCLA Sex Diary** (complete), IIEF, DISF-M-II | X        | X         | X         | X         | X   | X         |     |           |     |           |      |      |      |
| Vitality: SF-36-Vitality**           | X        | X         | X         | X         | X   | X         |     |           |     |           |      |      |      |
| Cognitive: BVRT, Card rotation, TMT  | X        |           |           |           |     |           |     |           |     |           |      |      |      |

V7.3 20130716 43
THE BONE TRIAL OF THE
THE TESTOSTERONE TRIAL

Sponsors
National Institute on Aging (NIA) and
Abbott Laboratories

NIH Grant Number
RO1AG037679-01A1

University of Pennsylvania Protocol Number
808676 (T Trial) and 814217 (Bone Trial)

Principal Investigator
Peter J. Snyder, MD

Study Drug Provider
Abbott Laboratories

IND Number
104707

Clinical Trials.gov Number
NCT00799617

<table>
<thead>
<tr>
<th>Bone Trial Original Protocol</th>
<th>July 29, 2011</th>
<th>Version Number: 1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Administrative Update</td>
<td>September 9, 2011</td>
<td>Version Number: 1.1</td>
</tr>
<tr>
<td>Administrative Update</td>
<td>January 30, 2012</td>
<td>Version Number: 1.2</td>
</tr>
<tr>
<td>Administrative Update</td>
<td>February 3, 2012</td>
<td>Version Number 1.3</td>
</tr>
<tr>
<td>Administrative Update</td>
<td>April 13, 2012</td>
<td>Version Number 1.4</td>
</tr>
<tr>
<td>Administrative Update</td>
<td>October 17, 2012</td>
<td>Version Number 1.5</td>
</tr>
</tbody>
</table>
# Table of Contents

A. Introduction .................................................................................................................. 5  
   1. Background .................................................................................................................. 5  

B. Specific Aims ............................................................................................................... 5  
   1. Study Aims .................................................................................................................. 5  
   2. Significance ................................................................................................................ 6  
   3. Innovation .................................................................................................................. 6  

C. Study Design ............................................................................................................. 6  
   1. Trial Subjects ............................................................................................................. 7  
   2. Inclusion Criteria ..................................................................................................... 7  
   3. Exclusion Criteria ................................................................................................... 7  
   4. Study Medication ................................................................................................... 8  
   5. Participating Study Sites ......................................................................................... 8  
   6. Reading Centers ..................................................................................................... 8  

D. Study Procedures ..................................................................................................... 9  
   1. Study Visits ............................................................................................................. 9  
   2. Quantitative Computerized Tomography (QCT) ..................................................... 9  
      a. Rationale ............................................................................................................. 9  
      b. Technique ......................................................................................................... 9  
      c. Quality Control ................................................................................................. 10  
      d. Bone Strength by Finite Element Analysis ....................................................... 10  
   3. Dual Energy X-Ray Absorptiometry (DXA) .......................................................... 12  
      a. Rationale .......................................................................................................... 12  
      b. Technique ......................................................................................................... 12  
      c. Quality Control ................................................................................................. 12  
   4. Calcium and Vitamin D Treatment ........................................................................ 13  
      a. Calcium ............................................................................................................. 13  
      b. Vitamin D .......................................................................................................... 13  
   5. Clinical Fractures ................................................................................................... 13  

E. Adverse Events ......................................................................................................... 13  
   1. General .................................................................................................................... 13  
   2. Additional Risks ................................................................................................... 14  
      a. Quantitative CT of Bone .................................................................................... 14  
      b. DXA Scan ......................................................................................................... 14  
   3. Protection Against Risk ........................................................................................ 15  

F. Statistical Considerations ......................................................................................... 15  
   1. Treatment allocation and balance ........................................................................ 15  
   2. Treatment blinding ................................................................................................. 16  
   3. Analytical methods and sample size estimations .................................................. 16  
      a. Analytical methods, sample size and power estimations for The Bone Trial .... 16
**Study Summary**

<table>
<thead>
<tr>
<th>Title</th>
<th>The Bone Trial of The Testosterone Trial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protocol Number</td>
<td>808676 (University of Pennsylvania, Data Coordinating Center)</td>
</tr>
<tr>
<td>Study Design</td>
<td>Randomized, placebo-controlled, double-blind</td>
</tr>
<tr>
<td>Study Duration</td>
<td>Five years</td>
</tr>
<tr>
<td>Study Centers</td>
<td>Nine clinical sites participating in the T Trial</td>
</tr>
<tr>
<td>Objectives</td>
<td>To test the hypothesis that testosterone treatment for one year will increase volumetric trabecular bone mineral density (vBMD) of the lumbar spine as measured by quantitative computed tomography (QCT) compared with placebo treatment</td>
</tr>
<tr>
<td>Number of Subjects</td>
<td>200</td>
</tr>
<tr>
<td>Diagnosis &amp; Main Inclusion Criteria</td>
<td>Participants in this study will be men enrolled in the T Trial, a randomized, placebo-controlled, double-blind study of seven coordinated trials in men &gt;65 years of age, using AndroGel or placebo gel for one year. Men in the trial have unequivocally low testosterone concentration (average of 2 morning testosterone values, &lt;275ng/dL), and symptoms and objective manifestations of mobility disability, low libido, or low vitality.</td>
</tr>
<tr>
<td>Study Product, Dose, Route, Regimen</td>
<td>AndroGel ® 1%, testosterone in an alcohol-water gel, administered transdermally in doses from 5 to 15 grams per day, adjusted as necessary to maintain the serum testosterone concentration within the range of normal for young men.</td>
</tr>
<tr>
<td>Duration of Administration</td>
<td>AndroGel or placebo will be administered for one year.</td>
</tr>
</tbody>
</table>
The Bone Trial of The Testosterone Trial

A. Introduction
This document is a protocol for a human research study. This study is 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.

The Bone Trial is being conducted as a study within The Testosterone Trial (TTrial). The TTrial is a randomized, placebo-controlled, double-blind study of seven coordinated trials in men greater the 65 years of age using testosterone gel or placebo gel for one year. The participants in this trial will be men presently enrolled in the TTrial who agree to also participate in the Bone Trial. As a result, the parameters of the TTrial protocol will be referenced throughout this description of the Bone Trial.

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, memory impairment, and osteoporosis. 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.

B. Specific Aims

1. Study Aims
The overall specific aim of the Bone Trial is to test the hypothesis that testosterone treatment for one year, compared with placebo treatment, of 200 men ≥65 years who have serum testosterone concentrations <275 ng/dL at Screening Visit 1 and < 300 ng/dL at Screening Visit 2, and an average serum testosterone concentration of < 275 ng/dL, and are participating in The Testosterone Trial, will improve bone quality.

The primary specific aim of the Bone Trial is to test the hypothesis that testosterone treatment for one year of elderly men with serum testosterone concentrations <275 ng/dL at 8AM will increase volumetric trabecular bone mineral density (vBMD) of the lumbar spine as measured by quantitative computed tomography (QCT) compared with placebo treatment. vBMD is the focus of the study because it is more responsive to testosterone than areal BMD (aBMD); because it separates the trabecular compartment from cortical shell; and because it avoids the artifacts of osteophytes and aortic calcification so common in elderly men.

The secondary specific aims are to test the hypotheses that testosterone treatment of these men will:
The Bone Trial of The Testosterone Trial

- Increase vbMD of the hip, as determined by QCT, because of the clinical importance of hip fractures and because this measure has been shown to predict hip fracture.
- Increase bone strength and strength-to-density ratio of the spine and hip, as determined by finite element analysis of the QCT data, because these parameters predict fractures better than bone density.
- Increase areal BMD of the lumbar spine and proximal hip, as measured by DXA. This assessment is proposed, despite its limitations, because it is the current standard clinical method of evaluating bone and because it does predict fracture risk.

Exploratory aims are to test the hypotheses that testosterone treatment will increase cortical bone in the spine and hip by QCT and reduce the incidence of radiographically confirmed clinical fractures. Another aim is to apply the fracture data from all 800 men in The Testosterone Trial for a sample size estimate for a larger, longer trial in which the effect of testosterone on the incidence of clinical fractures would be the primary end point.

2. **Significance**

The Bone Trial within the Testosterone Trial will be the first study of testosterone treatment in elderly hypogonadal men to employ QCT and thereby avoid the artifacts of DXA, but more importantly, the first to assess bone strength. Demonstration that testosterone treatment improves bone strength in elderly men would provide convincing scientific rationale for a larger, longer trial to determine if testosterone treatment reduces clinical fractures as well. The Bone Trial will also provide data on clinical fractures that will inform sample size estimation for a larger, longer trial.

3. **Innovation**

The Bone Trial is the first trial designed to adequately test the hypothesis that testosterone treatment of elderly hypogonadal men will improve their bone quality, because it would be the first to enroll elderly men who are unequivocally hypogonadal; the first to assess bone quality by a method most likely to detect an effect of testosterone in this population; and the first to assess bone strength in this population.

C. **Study Design**

The Bone Trial is a randomized, placebo-controlled trial to test the hypothesis that testosterone treatment for one year of men ≥65 years who have unequivocally low serum testosterone concentrations will improve their bone quality. Two hundred (200) men will be recruited for The Bone Trial from among 800 men who have been recruited, screened, and/or enrolled in The Testosterone Trial.

The Testosterone Trial has the following major features:
- Recruit men who are ≥65 years old and have unequivocally low early morning serum testosterone concentrations (average <275 ng/mL) on two days.
The Bone Trial of The Testosterone Trial

- Enroll 800 subjects in one or more of the three trials (Physical Function, Sexual Function, or Vitality) for which they qualify and consent to participate.
- Allocate men to treatment with testosterone or placebo gel double-blindly for one year; adjust the dose to maintain the serum testosterone within the reference range for young men while maintaining blinding.
- Exclude men who have diseases that testosterone could exacerbate and monitor those who do enroll for the development of these diseases.

The Bone Trial proposes the following additional features:

- Recruit 200 men from among 800 who have qualified for or are being screened to qualify for The Testosterone Trial.
- Exclude men who have conditions or are taking medications known to affect bone.
- Assess bone quality of these 200 men by QCT of the spine and hip, DXA of the spine and hip, and clinical fractures.

1. **Trial Subjects**

A total of 200 subjects will be recruited for The Bone Trial at 9 of the 12 Testosterone Trial sites. Subjects who are screened for The Testosterone Trial will be asked if they wish to participate in the Bone Trial sometime during their first screening visit of the Testosterone Trial. Those who do will undergo further screening for The Bone Trial. Please note that if each of the main trials (Vitality, Physical Function, and Sexual Function) within the Testosterone Trial has met its enrollment goals, subjects will be required to be eligible for and to participate in either the Bone Trial or the CV Trial, if they are still open to enrollment. At that time, all subjects who are screened for the Testosterone Trial will also be screened for the Bone Trial. In addition, all subjects who sign the T Trial baseline consent form will also receive details about the Bone Trial and will consent to the Bone Trial.

2. **Inclusion Criteria**

Interested subjects will be evaluated for eligibility for the Bone Trial. There will be no specific inclusion criteria for the Bone Trial. Participants will be required to sign an informed consent form specifically for the Bone Trial.

3. **Exclusion Criteria**

T Trial participants who have the following conditions or take the following medications will not be eligible to participate in the Bone Trial:

- Bone mineral density by DXA at the lumbar spine, total hip or femoral neck lower than -3.0.
- Elevated serum calcium (>10.5 mg/dL) at Screening Visit 1
- Medications that could influence bone, eg the following anticonvulsants: phenytoin, phenobarbital, carbamazepine, primadone, oxycarbazepine, topiramate, glucocorticoids (prednisone >20 mg/d >2 wk/year), bisphosphonates (alendronate, risedronate,
ibandronate), denosumab, and teriparatide. Calcium and OTC vitamin D supplements will be allowed.

- Any procedure or condition that prevents QCT analysis of the lumbar vertebrae

4. **Study Medication**

Testosterone (AndroGel III ®) or placebo (identical in appearance to AndroGel) 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 is 5.0 g once a day as described in the T Trial protocol. During the course of T Trial participation, the dose will be adjusted (higher or lower) to achieve a T level between 500 – 800 ng/dl.

Participation in the Bone Trial does not alter the dose or application procedures of the T or placebo gel for study subjects.

5. **Participating Study Sites**

The following nine (9) T Trial participating clinical sites will enroll men in the Bone Trial:

- Boston University, Boston, MA
- University of California at Los Angeles, Los Angeles, CA
- University of California at San Diego, San Diego, CA
- University of Alabama, Birmingham, AL
- University of Minnesota, Minneapolis, MN
- University of Pittsburgh, Pittsburgh, PA
- Yale University, New Haven, CT
- Baylor University, Houston, Texas
- Northwestern University, Chicago, Illinois

The University of Pennsylvania, Philadelphia, PA, will serve as the Data Coordinating Center (DCC) for the Bone Trial.

6. **Reading Centers**

The DXA reading center will be the SF Coordinating Center DXA Quality Assurance Group in the Department of Epidemiology and Biostatistics at the University of California at San Francisco.

The QCT reading center will be ON Diagnostics LLC, in Berkeley, CA.

Both reading centers will provide quality control for the respective procedures.
D. Study Procedures

1. Study Visits

Participants in the Bone Trial will follow the established T Trial visit schedule. Four (4) additional tests, some of which may be combined into a single visit, are required during the one year gel use phase: the first test at the time of the screening visit 2 or after, the second test before, during or after the baseline visit, and the third and fourth tests at the end of the treatment phase at the Month 12 visit.

- QCT of the spine and hip will be performed before, during or up to 2 weeks after the baseline visit but before the initiation of gel treatment and again after 12 months of treatment.
- DXA scan of the spine and hip will be performed at the SV2 visit or up to 2 weeks after the baseline visit but before the initiation of gel treatment and again after 12 months of treatment.

It is an important objective of the study to collect complete data on all subjects who enroll in the Bone Trial and sites will be asked to identify and approach men who are likely to be reliable subjects. However, a subject who enrolls in the Bone Trial may withdraw from the study without impacting his participation in the T Trial.

2. Quantitative Computerized Tomography (QCT)

a. Rationale

QCT will be the primary tool for assessing efficacy, because it can distinguish trabecular bone, which testosterone appears to affect primarily, from cortical bone. In addition, it is not artefactually influenced by osteophytes and aortic calcification, which are common in elderly men and confound the interpretation of DXA results. Finally, QCT data can be used for estimation of bone strength by Finite Element Analysis (FEA).

QCT of the spine will be the primary end point, because testosterone appears to have its greatest effect at this site. QCT of the hip will also be performed as a secondary end point, because of the clinical importance of hip fractures and because QCT of the hip predicts hip fractures.

b. Technique

Subjects will be instructed to wear metal free clothing and to remove any metal objects or jewelry from their body. They will be instructed to lie fully clothed on the scan table.

The QCT scans, covering the lumbar spine and pelvic region, will be acquired for each patient at 120 kvp. Subjects will be in the supine position in the scanner, arms above head, with an external mineral calibration phantom placed in the table cushion beneath. Reconstructions at 1mm thickness and spacing will be formed using a large field of view (FOV) and a standard (soft tissue type) kernal. The QCT exam will be completed in approximately 15 minutes. The same acquisition parameters will be used at all sites. Scans will be identified by Participant ID# and
initials only and sent via DICOM format directly to O.N. Diagnostics for BCT analysis. Quantitative measures of the QCT scans will be transmitted to the DCC in batches of 10.

All patient CT scans and radiographs will be reviewed by a staff radiologist at the clinical facility. All incidental findings will be made available to the patient or patient’s primary care physician, as appropriate.

c. Quality Control

Several steps will be taken to ensure consistency in data collection at the 9 imaging facilities. Each facility will undergo training to familiarize personnel with the study protocol. Site selection was influenced by the facility’s ability to provide reasonable assurance that the same CT scanner will be available for use with all patients at both time points. CT phantoms will rotate between sites every two months to be scanned using the same acquisition settings as for subjects. An external calibration phantom containing chambers of known bone mineral density concentration (Mindways, Model 3) will be used to convert each pixel value in the QCT scan into units of bone mineral density (mg/cm³). A torso phantom (Mindways, Model 3) will be used to correct for field nonuniformities due to beam hardening. These phantom scans will provide a means for cross-calibrating scanners to minimize machine differences across sites. If a scanner is changed during the study, additional phantom scans will be acquired before and after the change. O.N. Diagnostics will review the first five scans acquired at each facility, and every tenth scan thereafter, to identify and rectify any deviations from the protocol or loss in image quality.

d. Bone Strength by Finite Element Analysis

Several measures will be used to assess overall bone strength.

- Finite element models of the L1 vertebra will be constructed from digitized CT scans (1, 2). Each vertebral image (less posterior elements) will be segmented from the scan, rotated into a standard coordinate system, and resampled to 1 x 1 x 1 mm³ voxels. The CT number of each voxel will be converted into a bone mineral density value using the included external mineral phantom. The finite element mesh will be created by converting each voxel into an 8-noded brick element. The mean BMD across all voxels in the FE mesh will be used as the integral measure of volumetric density (vBMD).
- For each voxel, the axial elastic modulus ($E_z$) is calculated using the empirical correlation between elastic modulus and BMD for human vertebral trabecular bone (3). Elastic anisotropy of the bone is accounted for by assuming fixed ratios of the various elastic constants with respect to $E_z$(1).
- Material failure of the bone will be modeled by assigning an elastic-perfectly-plastic von Mises failure criterion. A thin layer of polymethylmethacrylate (PMMA) will be virtually placed at the ends of each vertebral model to create planoparallel surfaces on which uniform compressive displacement boundary conditions are applied.
- The compressive strength of the vertebra will then be computed as the total reaction force generated at 2% strain (applied displacement divided by bone height) in a non-linear stress analysis.
• The peripheral compartments are defined as all cortical bone plus any trabecular bone within 2 mm of the periosteal surface, and the trabecular compartment as all remaining trabecular bone. To assess the relative contribution to vertebral strength due only to the peripheral compartment, a fixed value of density is assigned to the trabecular compartment. Similarly, a fixed value of density will be assigned to the peripheral compartment to compute “trabecular strength.” To assess the relative contribution to vertebral strength due only to differences in external bone geometry, models will be created with a single uniform material property assigned throughout the entire vertebral body. In addition, vBMD of the cortical and trabecular bone combined (“integral” density), and the peripheral and trabecular compartments will be measured.

○ Finite element models will be constructed to estimate femoral strength (Figure 2)(4-6). The left proximal femur will be segmented from the calibrated CT images and resampled into 1.5 mm-sided voxels. Finite element mesh will be constructed by converting each voxel into an 8-noded brick element. Boundary conditions will be applied to simulate an unprotected fall to the side of the hip, the diaphysis angled at 15° with respect to the ground with 15° of internal rotation. To ensure consistent orientation, each bone will be registered to a reference bone in the fall orientation. The cortical and trabecular bone regions in these models will be distinguished from each other on the basis of apparent density (cut point of 1.0 g/cm³) and element-specific isotropic material properties will be derived from the calibrated volumetric BMD values using empirical relations, with different relations being used for the cortical and trabecular bone (7-9). From a nonlinear stress analyses femoral strength will calculated from the resulting force-deformation curve as the force at 4% deformation of the femoral head with respect to the greater trochanter.

Geometric strength, and strength and density of the peripheral and trabecular compartments, will be measured as described above for the vertebra with the exception that the peripheral compartment will consist of all cortical bone plus any trabecular bone within 3 mm of the periosteal surface.

Figure 2: Typical finite element model of the femur, showing 3D (A) and 2D sectional (B) views. The color-coding shows the spatial variation of material strength assigned to the individual finite elements.
3. **Dual Energy X-Ray Absorptiometry (DXA)**

   **a. Rationale**
   DXA of the spine and hip will be secondary end points, in spite of the limitations of DXA in elderly men, because DXA is the current standard method of assessing bone in clinical practice, because it does predict clinical fractures, and because previous studies of the effect of testosterone on bone suggested that testosterone treatment of elderly men should improve aBMD of the spine in men who have sufficiently low pretreatment serum testosterone concentrations.

   **b. Technique**
   Subjects will be asked to lie as still as possible on a padded table during the scans of the spine and hip. The detector is scanned over the area, generating images on a computer monitor. During the scan of the spine, the legs are supported to flatten the pelvis and lumbar spine. To scan the hip, the technologist will place the foot so that the hip rotates inward. The scan will take approximately 20 to 30 minutes to complete. Scans will be identified by Participant ID# and initials only. Analysis will be performed locally. Analyzed scans will be sent directly to UCSF for quality control and incorporation into the DXA dataset. Routine DXA scans will be transmitted to the DCC once a month. Baseline scans with a lumbar spine, total hip or femoral neck T score <3.0 and 12-month scans that show a <8% decrease in BMD will be sent immediately. Subjects will be given the report of their baseline scans by the clinical site. Subjects will not be given their 12-month scans unless the BMD decreases by 8%, confirmed by the DXA reading center.

   **c. Quality Control**
   All study sites will use a Hologic scanner to minimize site differences in the measurement of change in aBMD. All measures will be obtained and analyzed using Hologic specifications and a study-specific protocol. Quality control will be performed at the UCSF Coordinating Center, DXA QA Group (Dr. Ann Schwartz, PI) using procedures similar to those in the SOF, FIT, WHI, Health ABC, and PaTH studies. Densitometry operators will be certified in the study-specific protocol, and performance will be monitored by central review of selected scans. A complete scan database will be maintained at the CC. Sites will report possible cases of low BMD at baseline or the 12-month visit (BMD T-score < -3.0 at the total hip, femoral neck, or lumbar spine) to UCSF for confirmation. Sites will also report possible cases of excessive bone loss (change from baseline to 12-months >8%) to the DXA QA center for review. If confirmed, UCSF will notify the clinic coordinator who will notify the participant of low BMD or excessive bone loss.

   Densitometer performance will be monitored with longitudinal spine phantom scans. Measures of bone loss will be derived from analyses of paired scans using the 'compare' feature, so that baseline and follow-up scans are analyzed with the same software version. However, to the greatest extent possible, changes in scanner software will be avoided during the study.
4. **Calcium and Vitamin D Treatment**

Many older men have inadequate intake of calcium and vitamin D. Therefore, participation in the Bone Trial includes providing calcium and OTC vitamin D supplements to all men in the Bone Trial. It is likely that this supplementation will improve bone density in the treatment as well as the placebo group. As a result, the sample size estimation has been calculated based on a study that employed similar supplementation. Subjects will be reminded to take calcium and vitamin D at each study visit.

   a. **Calcium**

   Subjects will be given a supply of calcium supplements and instructed to take 1200 mg of elemental calcium a day, which is the amount recommended by the NIH and the National Osteoporosis Foundation.

   b. **Vitamin D**

   Subjects will be given a supply of OTC vitamin D and instructed to take 800 units of vitamin D a day.

5. **Clinical Fractures**

Although the study will not likely have enough subjects followed for a sufficiently long time to have adequate power to detect a difference between treatment groups in the incidence of fractures, clinical fractures in all 800 T Trial subjects will be tracked and evaluated for trends. The same procedures for fracture ascertainment and adjudication will be followed as in the MrOS study, which are:

   - Information on fracture history will be obtained prior to randomization
   - After randomization, each trial site will inquire specifically about fractures at each contact
   - If informed of a fracture, the trial site will inquire about the circumstances (e.g. fall, motor vehicle accident, etc.), anatomic site of fracture, and if and where x-rays were taken; obtain the radiology report and other clinical documentation, such as operative and orthopedic notes, etc.
   - Clinical sites will send reports and other clinical documentation films directly to the UCSF Reading Center, where Dr. Douglas Bauer will compile the data. Fractures associated with major trauma, e.g., motor vehicle accident, may be excluded from analysis.

E. **Adverse Events**

   1. **General**

   All aspects of recording and reporting adverse events described in the T Trial protocol apply to the Bone Trial. At each Bone Trial visit, adverse event data will be collected and reported in the same manner as in the T Trial with the identical oversight and safeguards.
2. **Additional Risks**

The additional risks associated with the Bone Trial participation are related to the additional radiographic tests.

a. **Quantitative CT of Bone**

Radiation exposure from CT scans introduces a potential risk to study participants. The effective dose to each subject is expected to be approximately 1.8 mSv during the standard lumbar CT exam protocol and 2.4 mSv during the pelvic CT exam protocol. Assuming an annual natural background radiation dose of 2.4 mSv, the total per patient dose for this study (~4.2 mSv) has the same relative risk as about 1¾ years of natural background radiation. The total radiation exposure for the study will be less than typical exam doses (Table 1) and well below the maximal acceptable dose for normal subjects. The subjects will be counseled about the amount of radiation they will receive as a result of participation in the study as part of Informed Consent procedures.

The QCT test of the spine and hip will expose the subject to a dose of 4.2 mSV of additional radiation.

<table>
<thead>
<tr>
<th>Exam</th>
<th>Dose (mSv)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colonography *</td>
<td>2.4-7.8</td>
</tr>
<tr>
<td>Body CTA *</td>
<td>6.1</td>
</tr>
<tr>
<td>Routine Liver *</td>
<td>8.0</td>
</tr>
<tr>
<td>Coronary CTA *</td>
<td>14.5</td>
</tr>
<tr>
<td>Proposed QCT Spine &amp; Hip</td>
<td>4.2</td>
</tr>
</tbody>
</table>


b. **DXA Scan**

DXA of the spine and hip will expose subjects to a very low dose of radiation, 2.2 microSv for the lumbar spine scan and 5.1 microSv for the hip scan, which is less than that of a standard chest x-ray.

**Time Burden**

The tests proposed in the Bone Trial will not likely be burdensome by themselves, but because the men who will participate in the Bone Trial will also be participating, and have a large number
of tests in the overall Testosterone Trial, the total subject burden could be considerable. The site staff will be instructed to be cognizant that some subjects could have difficulty participating in The Bone Trial tests, as well as the overall Testosterone Trial, and to consider spreading the testing over more than one day.

3. **Protection Against Risk**

Care is taken during QCT and DXA imaging tests to use the lowest radiation dose possible while producing the best images for evaluation. National and international radiology protection councils continually review and update the technique standards used by radiology professionals. State-of-the-art x-ray systems have tightly controlled x-ray beams with significant filtration and dose control methods to minimize stray or scatter radiation. This ensures that those parts of a participant's body not being imaged receive minimal radiation exposure.

The clinical sites participating in the Bone Trial have been chosen for their use of state-of-the-art scanning equipment and techniques and for their imaging experience in many similar trials.

F. **Statistical Considerations**

1. **Treatment allocation and balance**

The standard approach to treatment allocation is randomization, balancing on prognostic factors by stratification, but this approach is not practical for The Testosterone Trial is actually several trials in one, which would greatly complicate balancing by stratified randomization, since there would be more than a thousand strata. Stratifying using fewer variables would risk imbalance with regard to one or more key prognostic factors. The minimization has been selected as a more suitable alternative in this setting. Minimization is a covariate-adaptive procedure that can be employed even when there are a large number of factors to be balanced (10, 11), because balance is achieved within strata defined by the sum of categories for the balancing factors rather than the product, as with stratified randomization. Minimization is based on real-time calculations of treatment assignment, balanced with respect to each important prognostic variable. The subject is assigned to the treatment that produces the best balance overall, considering all relevant factors, with a specified high probability. Because the treatment assignments generated by this method are no more predictable than they would be with stratified randomization, the assignments can be considered essentially random.

Minimization will be used with a random component, for treatment allocation, assigning subjects to the optimal balancing treatment with probability 80%. Factors for balancing will include a dichotomous variable for each of the primary efficacy trials in which a subject may participate, study site, baseline testosterone concentration, age, and current use of an antidepressant and PDE-5 inhibitor; balancing factors will be accounted for as covariates in all analyses. Using this approach, simulation studies have performed simulations that verify that it provides the desired balance with regard to the prognostic variables planned for use.
2. **Treatment blinding**

Several methods are used to maintain blinding. Treatment assignment will be known only to the data coordinating center and research pharmacy. The testosterone and placebo preparations will look, smell, and feel the same. Maintenance of blinding when the testosterone dose is adjusted is achieved by adjusting the dose in a participant using placebo gel. Clinical sites will not have access to the testosterone-dependent laboratory tests, e.g. hemoglobin or PSA.

3. **Analytical methods and sample size estimations**

Each of the individual efficacy trials in The Testosterone Trial, such as The Bone Trial, is considered a separate trial, so the results will be analyzed separately. Data in The Bone Trial will be analyzed as continuous variables. All analyses will incorporate accounting for balancing covariates.

a. **Analytical methods, sample size and power estimations for The Bone Trial**

The primary end point for The Bone Trial will be change from baseline to 12 months in trabecular volumetric bone mineral density (vBMD) in the spine as measured by quantitative computed tomography (QCT). The analytic approach will be comparison of treatment groups using ANCOVA, accounting for balancing factors, the same as used for balancing in treatment allocation. Data from a prior study in hypogonadal men showed an increase in mean trabecular vBMD of 14% after 18 months of testosterone treatment, with a SEM of 3% (12). The testosterone levels of men in that study were similar to those in the proposed study. It is conservatively assumed that improvement will be linear and will be 9% by 12 months, with the same SEM. To detect such an effect with 90% power will require 86 subjects per arm or 172 subjects for both arms. An additional 15% accrual will compensate for dropout and noncompliance, leading to a total sample size of 200.

Secondary end points will include trabecular vBMD of the hip by QCT and bone strength and strength-to-density ratio by finite element analysis in the spine and hip by QCT. The power to detect differences between the AndroGel and the placebo groups in these end points is based on data from a study of ibandronate versus placebo treatment of postmenopausal women for one year (5). Assuming a sample size of 172, this study has 90% power to detect a similar difference between treatment groups in hip trabecular vBMD, 98% power in hip trabecular strength, >99% power in hip strength/density ratio, and 90% in spine trabecular strength.

Areal bone mineral density of the lumbar spine and hip by DXA will also be secondary end points. Data from two prior trials (13, 14) suggest a threefold improvement in aBMD of the spine compared with a placebo arm, given a pretreatment serum testosterone <275 ng/dL. Given a sample size of 172, there should be 90% power to detect a doubling (a more conservative estimate) between the two groups.

Exploratory end points will be changes in cortical bone in the spine and hip by QCT and clinical fractures. Serum markers of bone turnover tests will be conducted at a later date if additional funding becomes available. Clinical fractures will be assessed in all 800 subjects in The Testosterone Trial. Even if testosterone treatment should be associated with a reduction in...
clinical fractures, the mechanism could be an improvement in muscle strength and reduced
tendency to fall. For this reason, and because falls is an endpoint in all 800 men, finding fewer
falls in the testosterone group than the placebo group might be predictive of a lower fracture
rate. Regression analyses will be performed to assess whether certain baseline characteristics,
including baseline vBMD and aBMD values and age, are associated with greater change in
trabecular vBMD of the spine at 12 months.

Another exploratory objective is development of a sample size estimate for a longer, larger trial
in which the effect of testosterone on clinical fracture incidence in elderly men is the primary
end point. The design of such a trial requires data on the expected fracture rate in untreated
men and an estimate of the potential treatment effect in order to develop sample size
requirements. The entire 800 men in The Testosterone Trial treated for one year and followed
for two years will provide data on fracture incidence. The T Trial results will be compared to
those from the MrOS study, in which clinical fractures occurred at a rate of 19/1000 person-
years (Cauley, personal communication). If clinical fracture incidence in The Testosterone Trial is
similar to that in MrOS, weighted according to expected age distribution, 25-35 fractures may be
observed.

The Bone Trial will also provide an estimate of the effect of testosterone treatment on fracture
incidence. Although the number of fractures in the two treatment groups will be too small to
draw firm conclusions, any trend will be of interest. The association between bone strength by
finite element analysis (FEA) and clinical fracture data will be determined, because of the
findings from MrOS of a strong association between bone strength by FEA and incidence of hip
fracture. If the T Trial fractures data confirm those of MrOS, and also predict that the
testosterone group would have fewer fractures than the placebo group, these results would
provide both a rationale and a quantitative basis for conducting a larger, longer study with the
incidence of clinical fractures the primary outcome.

4. Managing Missing Values
The primary analysis will include only men who have measurements at both baseline and 12
months; however, a variety of sensitivity analyses will be performed to assess the potential
impact of missing 12-month outcomes. Baseline characteristics of men who do and who do not
have a 12-month measurement will be compared to help determine any possible biases
attributable to the pattern of missing values.

5. Intent-to-Treat
Primary analyses will follow the “intent-to-treat” (ITT) principle; individuals will be analyzed
according to their assigned treatment group whether or not they continue treatment. Every
attempt will be made to follow and evaluate all enrolled subjects.

6. Multiple analyses
Because a large number of analyses are likely to be conducted – e.g., individual domains
contributing to the instrument scores, alternative statistical approaches, and interest in
particular subsets will lead to multiple analyses—the likelihood of multiplicity-induced false positive outcomes will be explored. The approach to multiple comparisons will be based on controlling the false discovery rate according to the methods of Benjamini and Hochberg.

G. Study Management
The Data Coordinating Center (DCC) at the University of Pennsylvania will develop a secure electronic data management system for the collection, storage, and management of the Bone Trial data. Lab data will be transferred from Quest using the secure server. The DCC will follow all of the Standard Operating Procedures (SOP) that have been established in the T Trial to achieve the same level of data integrity and security.

1. Data Management System Components
The Data Coordinating Center (DCC) at the University of Pennsylvania will develop an electronic data management system module for the collection, storage, and management of Bone Trial data, as consistent with the main Testosterone Trial Protocol.

Manual of Procedures (MOP) – A MOP will describe the technique for conducting QCT and DXA scans on participants. This manual will provide detailed information for testers in the preparation, acquisition, transfer, and data quality assessment of images acquired in the Bone Trial. The MOP will also provide instruction in case report form completion, use of the electronic data management system. Review of this manual will be part of the initial training conducted by the reading center.

Training and certification procedures – The DXA and QCT Reading Center Directors will conduct training in these respective procedures. Each DXA and QCT technologist involved in the Trial should have a complete understanding of the protocol and experience in acquiring images and operating the scanner. To ensure quality control, each site should designate specific technicians to perform the scans. The Trial will provide training in study-specific procedures, but operators are assumed to be proficient in image acquisition.

DXA training will be supervised by Dr. Ann Schwartz, the Director of the DXA Reading Center, who will prepare a manual of procedures for DXA. She will conduct training of DXA technicians using a standard PowerPoint presentation and teleconference system. Using the standard interactive presentation and teleconference, each site can get specific training and follow-up as needed. Further, quality assurance will be performed on scan (see below), and necessary feedback given to the specific technologist promptly after the scan is received at the Reading Center. This will allow for rapid identification of improper scan techniques and has been shown to dramatically decrease inadequate data sets.

QCT training will be conducted by Dr. David Kopperdahl of ON Diagnostics. He will prepare a manual of procedures for QCT and visit each site at least once. During these visits he will include in the training a supervisor as well as the technicians who will perform the scans, so if the technicians who are trained leave, someone who knows the protocol will be able to help train a new technician.
The DCC will establish a data structure and transfer process for each reading center and test these procedures at the start of the trial. The DCC will monitor quality control of data received from the QCT and DXA reading centers. A data validation plan, rule set specifications, and programming logic to implement data validation rules will be applied to the data.

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.

3. **Data Security**

The data management system will be the same system as the main Testosterone Trial Study and is 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.

a. **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.

b. **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.
H. Reference List


VII. Bone Trial

VII.A Primary Endpoint

VII.A.1 Primary Analysis. The primary analysis will assess differences in the percent change from baseline to month 12 volumetric bone mineral density (vBMD) of trabecular bone of the average of L1 and L2 (spine) as measured by QCT in an intent-to-treat comparison of all randomized subjects qualifying for the Bone Trial. Only men with a baseline and 12-month trabecular vBMD measurement will be included in the primary analysis. A multivariable linear model will be used to evaluate the effect of Androgel while adjusting for baseline trabecular vBMD and balancing factors (study site, indicators of participation in each of the primary efficacy trials, baseline testosterone concentration (<200), age (≤ 75), use of anti-depressants, and use of PDE-inhibitors). The null hypothesis of no effect of Androgel on 12-month change in trabecular bone density will be evaluated through the two-sided Wald test and confidence intervals of the Androgel coefficient.

VII.A.2 Sensitivity Analysis for Missing Data. Sensitivity analyses for missing post-baseline measurements will be conducted using the methods described in Section V.A.2.

VII.B Secondary endpoints – Analysis of all secondary endpoints will use the same method as the primary endpoint analysis, adjusting for the corresponding baseline measurement of the outcome of interest. The outcome will be the percent change in the measure from baseline unless otherwise indicated.

QCT measures

VII.B.1.1 Change in average trabecular vBMD of L1 and L2. The impact of testosterone on change will be assessed through a multivariate linear model adjusting for balancing factors as detailed in VII.A.1. This analysis differs from the primary analysis in that the change outcome is not standardized by the baseline value. The baseline value will still enter the model as an adjustment variable.

VII.B.1.2 Average trabecular bone strength of L1 and L2 (spine)
VII.B.1.3 Average cortical bone strength of L1 and L2 (spine)
VII.B.1.4 Average whole bone strength of L1 and L2 (spine)
VII.B.1.5 Average trabecular bone vBMD of the left and right hips
VII.B.1.6 Average whole bone strength of the left and right hips

DXA measures

VII.B.1.7 Average aBMD of L1 and L2 (spine)

VII.B.2 Multiplicity. Adjustment for multiple comparisons will be considered in the interpretation of statistical significance. The total number of secondary endpoints will be presented, and the
number of significant tests will be interpreted relative to the number of significant findings expected under the chosen type I error rate (α=0.05).

**VII.C Exploratory endpoints.** Analysis of all exploratory endpoints will use the same method as the primary endpoint analysis, adjusting for the corresponding baseline measurement of the outcome of interest. The outcome will be the percent change in the measure from baseline to 12 months unless otherwise indicated.

**QCT**

- VII.C.1.1 Average cortical bone vBMD of L1 and L2 (spine)
- VII.C.1.2 Average whole bone vBMD of L1 and L2 (spine)
- VII.C.1.3 Average geometric strength of L1 and L2 (spine)
- VII.C.1.4 Ratio of average L1 and L2 strength to average L1 and L2 vBMD ratio
- VII.C.1.5 Trabecular vBMD of the spine. The primary endpoint analysis of the percent change in average trabecular vBMD of L1 and L2 (spine) will be repeated with BMI included as an additional adjustment variable.
- VII.C.1.6 Average cortical bone vBMD of the left and right hips
- VII.C.1.7 Average whole bone vBMD of the left and right hips
- VII.C.1.8 Average trabecular bone strength of the left and right hips
- VII.C.1.9 Average cortical bone strength of the left and right hips
- VII.C.1.10 Average geometric strength of the left and right hips
- VII.C.1.11 Ratio of average left and right hips strength to average left and right hip vBMD ratio

**DXA**

- VII.C.1.12 Average abM of the left and right hips (entire proximal femur)
- VII.C.1.13 Average aBMD of the left and right hips (femoral neck)
- VII.C.1.14 Clinical Fractures. The number of fractures by arm and site (e.g. hip, ankle, etc.) will be reported. If there are enough fractures to allow comparisons, differences in the experience of clinical fractures across all trials for the Androgel versus placebo arm will be evaluated by logistic regression.

Fracture experience will be dichotomized to reflect ever experiencing versus never experiencing a fracture. The model will adjust for balancing factors and efficacy trial participation.

**VII.C.2 Exploratory Analyses.** Potential predictors of the percent change in the following measures will be evaluated through multivariable linear regression. Baseline characteristics, such as vBMD, aBMD, bone strength, and age, will be evaluated as potential predictors, adjusting for treatment and balancing factors.
VII.C.2.1 Average trabecular vBMD of L1 and L2 (spine)
VII.C.2.2 Average cortical vBMD of L1 and L2 (spine)
VII.C.2.3 Average whole bone vBMD of L1 and L2
VII.C.2.4 Average trabecular vBMD of the left and right hips
VII.C.2.5 Average cortical vBMD of the left and right hips
VII.C.2.6 Average whole bone VBMD of the left and right hips
VII.C.2.7 Data-reduced composite bone endpoint. Factor analysis will be used to reduce the dimension of all bone variables to a single bone outcome. Composite bone measures for given categories (density or strength) will be computed similarly and compared between treatment arms.

VII.D Subgroup Analyses. Subgroup analyses will be completed by repeating the primary analysis with interactions between treatment and subgroup-defining variables as listed below.

VII.D.1a DXA (femoral neck) above and below -2.0. If the percentage of subjects falling below -2.0 on their baseline femoral DXA is too small, subgroups will be defined according to those that fall above and below the median femoral DXA value.
VII.D.1b Free testosterone
VII.D.1c Age <75
VII.D.1d FRAX score (may be modified according to available variables)

VII.D.2 Baseline Testosterone and Estradiol as Predictors of Response. For primary, secondary, and exploratory outcomes showing a significant (p<0.05) effect of treatment, the analysis will be repeated including interactions of assignment to treatment with baseline free and total testosterone and baseline estradiol. In separate models, continuous and binary free and total testosterone and estradiol interactions with treatment will be considered.

VII.E Mediation analyses. Whether the effect of testosterone is mediated by given factors will be evaluated by the product of coefficients approach. Mediation analysis of the testosterone effect on bone will consider month 6 levels of the potential mediator.

VII.E.1 Estradiol
VII.E.2 DHT (Androgen receptors)