

**Analysis Plan for the primary results of the
Testosterone and Exercise in Aging Men (TEAM)
AG-02-1056
(DRAFT: 13 December 2010)**

Note: this analysis plan refers to low and high dose testosterone groups. The term “dose” is used for convenience; it represents low concentration dosing and high concentration dosing as described in protocol section 6b.

1. Trial structure

1.1 Hypotheses and specific aims (from protocol):

Specific Aim 1: To determine the effects of 1 year of low-dose (25 mg/d; 2.5G as AndroGel™) T versus usual-dose (50 mg/d; 5.0G as AndroGel™) T treatment on body composition, strength, power, endurance, and functional performance, as well as on hematocrit, prostate size/symptoms (and PSA) and symptoms of obstructive sleep apnea (OSA) in healthy older men with low-normal to slightly below normal serum T levels.

Hypothesis 1 (Main effect of T): Long-term (1 year) T treatment in older men using a lower than usual replacement dose will have beneficial effects on *body composition, strength, power, and physical functional performance*.

Secondary Hypothesis: Low-dose T supplementation (25 mg/d; target serum concentration of 400-550 ng/dL) will produce fewer adverse effects compared to usual replacement dose T (50 mg/d; target serum concentration of 600-1000 ng/mL).

Specific Aim 2: To determine the effects of 1 year of low vs. usual-dose T supplementation with and without a standard progressive resistance training (PRT) program on body composition, strength, power, endurance and physical functional performance, as well as on hematocrit, prostate size/symptoms and symptoms of obstructive sleep apnea (OSA) in healthy older men with low-normal to slightly below normal serum T levels.

Hypothesis 2 (Interaction between T and PRT): PRT will enhance the anabolic effect of T treatment in older men, producing similar (or greater) positive anabolic effects on body composition, strength, power and physical functional performance at lower T doses. We hypothesize that low dose T + PRT will have effects greater than low-dose T or (usual) replacement dose T alone, but not as great as usual dose T + PRT (usual T+PRT > low T+PRT > usual T > low T).

1.2 Trial structure (abstracted from protocol): The study was designed as a one-year, prospective, randomized, double-blind, placebo-controlled, complete, balanced, 2-way factorial design in healthy, community-dwelling older (≥60 yrs) men with low-normal to slightly below normal serum total T levels. The 2 factors were T supplementation (placebo, 25mg/d, 50mg/d) crossed with PRT (3x/wk versus none). The control exercise group was offered PRT after completion of the study (wait list control). The objective was to randomize 150 men to 6 treatment groups

Table 1

Testosterone Dose (steady state serum T range)

	0mg/d (200-350ng/dL)	25mg/d (400-550 ng/dL)	50mg/d (600- 1000ng/dL)
No PRT (wait list)	Group 1 (n=25)	Group 2 (n=25)	Group 3 (n=25)
PRT (3x/wk)	Group 4 (n=25)	Group 5 (n=25)	Group 6 (n=25)

The staff and investigators (except the research pharmacist, the Data/Safety Monitoring Board and data manager) were blinded to the drug treatment regimens. The staff and investigators were blinded to exercise assignment (except for the nurse coordinator and the exercise trainers). Subjects were blinded to the drug treatment.

As stated in the protocol (section 3a), the hypotheses will be tested as follows:

- **Hypothesis 1: (T main effect)**-- the potential beneficial effects of 12 months of the low-dose T supplementation (25 mg/d, 400-550 ng/dL; **group 2**) compared to usual replacement dose T (50 mg/d, 600-1000mg/dL; **group 3**) and placebo (0 mg/d, 200-350 ng/dL; **group 1**) on body composition, strength/power and functional performance.
- **Hypothesis 2: (Interaction between T and PRT)**-- the effects of low-dose T + PRT (**group 5**) compared to usual dose T + PRT (**group 6**) and PRT alone (**group 4**) on body composition, strength/power and functional performance.
- We will also compare the effects of the T gel only groups (**groups 1, 2 and 3**) with PRT alone (**group 4**) and evaluate possible interactions between T supplementation and PRT.

We will recruit enough subjects to permit a 15-20% dropout rate, and still complete the necessary 25 subjects per group.

1.3 Sample size (protocol section 13g):

We will recruit a total of 180 subjects for this study. We expect this to yield at least 25 per group after dropouts, which give us $\geq 90\%$ power to detect the main effects of exercise and T use on the primary outcomes of strength, function, and composition. Power calculations were carried out using the method of Muller, which allows calculating power for the same tests of interaction and main effect proposed for the analyses.

1.3 Key outcome measures (at baseline, 6 months, and 12 months):

Please define the following (give units); organize into categories of (1) composition, (2) strength, and (3) physical function:

csPFP:

Power Rig:

1RM:

Body Composition DXA/CT (are there more than one measure?):

BMI:

Biodex:

VO2max:

2. Analysis Plan

2.1 Protocol analysis plan (section 13f):

Structuring the analyses: The trial is a 2x3 factorial design (6 treatment groups) with 3 main categories of outcomes (body composition, strength, physical function) measured at 2 time points (6 and 12 months). The number of possible analyses is very large. In order to avoid data-driven conclusions and control the risk of a false-positive trial (type I error rate), all analyses will be pre-specified and prioritized in a written statistical analysis plan (SAP), which will be prepared prior to any data analysis. The SAP will follow the general plan outlined above (section 3a). This approach directly addresses the trial hypotheses, and avoids data-driven decisions in the analyses. The two testosterone

treatment groups will be combined in a primary analysis to test outcome differences between placebo and this combined T group. The a-priori hypothesis is that low testosterone will be as effective as high testosterone, but will have fewer side effects. As in section 3a, separate analyses will be conducted in the no-exercise (group 1 versus combined groups 2&3) and exercise (group 4 versus combined groups 5&6) arms.

Primary analysis: The effect of testosterone in combination with exercise (group 4 vs 5&6) will be considered the primary comparison. The effect of testosterone without exercise (group 1 vs 2&3) will be considered secondary. The prioritization is adopted because the a-priori hypothesis is that the largest effects on physical function will be achieved with exercise plus testosterone.

The primary question regards the effect of testosterone and exercise on physical function at 52 weeks. This question will be tested in a single statistical test as follows:

2-sample t-test of the difference in mean csPFP with placebo plus exercise (group 4) versus the combined testosterone plus exercise groups (5&6). This comparison will be adjusted for baseline csPFP in an analysis of covariance in order to increase precision (gain power). The following model will be fit:

$$csPFP_{52} = \beta_0 + \beta_1 Tx + \beta_2 csPFP_0$$

where $csPFP_{52}$ and $csPFP_0$ denote physical function score at baseline and 52 weeks, and Tx denotes treatment group (0 for placebo and 1 for testosterone). The overall conclusion for the effect of testosterone on physical function (the primary objective of the TEAM trial) will be judged by the statistical significance of the β_1 coefficient using a Wald test. A one-sided p-value less than 0.025 will define statistical significance.

Secondary Analyses: The primary conclusion for the trial will be based on the primary analysis. Key secondary questions will also be analyzed; these include:

- Are the physical function effects with exercise the same without exercise (compare groups 1 vs 2&3)?
- Is the side effect profile affiliated with low-dose testosterone lower than that of high-dose?
- Is the strength/physical function profile different between low and high-dose testosterone?

In addition to these key secondary questions, secondary analyses will also evaluate whether the primary conclusions are supported by other measures of treatment effect; specifically:

- Do other measures of physical function give the same conclusion as csPFP?
- Are the 26-week results consistent with the 52-week results?
- Are the effects on strength and body composition consistent with the effects on physical function?

2.2 Preliminary evaluation of data:

Purpose: To identify potential data entry errors and cases that might be influential in the primary analyses.

Methods: (Combined intervention groups)

- Tabulation of descriptive statistics for key variables (including missing observations).
- Amount of missing data by variable.

- Note: formal treatment of missing data in the statistical analyses is discussed in section 3.1.

2.3 Description of study process

Purpose: To compare how the study unfolded to determine if changes in analysis plans are warranted.

Methods/Approach:

Study flow figure

See if the following are already in Tammy's figure:

Reasons for screen failure

Reasons for refusal

Participants randomized (number to each group)

Retention by group

Number subjects in analysis data set

2.4 Description of participants and treatment

Purpose: To summarize the characteristics of study participants in order to (1) inform generalizability and (2) assure comparability of intervention groups. Second, to describe the actual treatment received

Methods/Approach: Construct "Table 1" describing the baseline characteristics of each intervention group.

Each table summarizes:

Demographics comparing completers vs LFUs (are dropouts different from completers in age, race, education, drug class use, etc)

Baseline values of primary and secondary outcome variables

Preliminary versions of these tables are attached as tables 1, below.

2.5 Description of treatments received

Purpose: To describe the level of testosterone received and the serum levels achieved in the different treatment groups.

Methods/Approach: [NOTE: create table shells.] Table 2a: mean testosterone prescription dose, MPR, and mean serum levels for the 3 testosterone dose levels. Table 2b: exercise attendance/compliance for the exercise (and no-exercise?) groups.

ACTION: Review/modify content of table 2a-b.

2.6 Primary results

Purpose: To present the primary results for the **intention to treat (ITT)** analysis.

Structuring the analyses: The trial is a 2x3 factorial design (6 treatment groups) with 3 main categories of outcomes (body composition, strength, physical function) measured at 2 time points (6 and 12 months). The number of possible analyses is very large. This section pre-specifies the analytic strategy in order to both document and control the risk of false positive conclusions (type I error rate).

Primary analyses: The primary question regards the effect of testosterone and exercise on physical function at 52 weeks. There is a single statistical test for this question:

- 2-sample t-test of the difference in mean csPFP with placebo plus exercise (group 4) versus the combined testosterone plus exercise groups (5&6). This comparison will be adjusted for baseline csPFP in an analysis of covariance in order to increase precision (gain power). The following model will be fit:

$$csPFP_{52} = \beta_0 + \beta_1 Tx + \beta_2 csPFP_0$$

where $csPFP_{52}$ and $csPFP_0$ denote physical function score at baseline and 52 weeks, and Tx denotes treatment group (0 for placebo and 1 for testosterone). The overall conclusion for the effect of testosterone on physical function (the primary objective of the TEAM trial) will be judged by the statistical significance of the β_1 coefficient using a Wald test. A one-sided p-value less than 0.025 will define statistical significance.

As described below, several secondary analyses are planned to evaluate mechanism related to either a positive or negative primary analysis.

ACTION: Review/revise draft tables 3a-3d (attached): How should we present the primary analysis of the combined T-groups? What variables should be included in these tables – consider subscales of the PFP along with other measures of physical function (SF-36?), of strength, or of body composition.

2.7 Secondary analyses

ACTION: Review/revise content of table 3.

Purpose:

1. To evaluate *coherence*; i.e., are the primary conclusions supported by other measures of treatment effect? Specifically:
 - Do other measures of physical function give the same conclusion as csPFP?
 - Are the 26-week results consistent with the 52-week results?
 - Are the effects on strength and body composition consistent with the effects on physical function?

Methods/Approach: Evaluate the results shown in table 3 to answer the following:

1. Do other measures of physical function give the same conclusion?
 - a. Redo primary analysis using subscales of csPFP.
2. Are 26-week results consistent with the 52-week results?
 - a. Redo the primary analysis using 26-week csPFP measurements.
3. Are the results on strength consistent with the physical function results?
 - a. Repeat primary analysis using 52-week measures of 1RM (and other strength measures).
 - b. Repeat primary analysis using 26-week measures of 1RM (and other strength measures).
4. Are the results on composition consistent with the strength and physical function results?
 - a. Repeat primary analysis using 52-week measures of DXA (and MRI composition measures).
 - b. Repeat primary analysis using 26-week measures of DXA (and composition measures).

2.8 Adverse effects

ACTION: Bob needs to develop adverse event categories (and how they are defined). How are the following scored? How should these results be organized in a table?

Purpose: To evaluate potential adverse intervention effects.

Methods/Approach: The analysis of table 3a will be repeated for measures of adverse effects including:

1. SAE (including hospitalizations)
2. HCT
3. LFT
4. OSA / Somnolence
5. PSA
6. AUA
7. Cardiovascular events (developing hypertension, renal function, edema)

ACTION: Further analyses: complete the analysis plan to here. Keep a change log of subsequent analyses and a dated list of when analyses were completed. This will allow us to demonstrate what was pre-specified.

3.0 Other analyses:

3.1 Tertiary analyses:

1. Is there any evidence of differences between T-dose? (different paper?)
2. Do effects differ by serum T? (different paper?)
3. Robustness of conclusions to confounders?
4. Other data driven analyses (identified as such).
5. Compliers analysis?

3.2 Subgroup analyses

Purpose: (1) To evaluate the consistency of the primary conclusions across important subgroups; (2) to identify subgroups (if any) that differed in their response.

Methods/Approach: The primary results (primary outcome and intervention target measures) will be re-analyzed in the following subgroups:

1. _____
2. _____

3.3 Satisfaction/Compliance (intervention group only?):

Purpose: To present a preliminary overview of participant satisfaction and compliance with key elements of the intervention.

Methods/Approach: (Note: the specific measures of compliance and satisfaction have not yet been defined.)

1. Compare compliance among subgroups (T-levels and exercise).
2. Evaluate satisfaction with key elements of the intervention.
3. Evaluate prior T use assumption.

4. Robustness of the conclusions

4.1 Missing data

4.2 Influential cases (are our results due to a single individual)???

4.3 Potential confounders: if any of the following variables show important baseline imbalances between treatment groups, then an adjusted regression analysis will be conducted to determine if the primary conclusion is substantively altered.

Age

Baseline testosterone

BMI

Lean body mass

Table 1a Baseline demographic characteristics of study participants by randomization group

	Exercise Group N=			No Exercise Group N=		
	Placebo (n=)	Low Range (n=)	High Range (n=)	Placebo (n=)	Low Range (n=)	High Range (n=)
Ethnicity – no. (%)						
Hispanic						
Non-Hispanic White						
Other						
Race						
Caucasian						
African American						
Asian						
Native American						
Other						
Age – yr. (std)	92.3 (15.5)					
Education (yr)						
Blood pressure - mm Hg (std)						
Systolic						
Diastolic						

Table 1b Baseline biochemical characteristics of study participants by randomization group						
	Exercise Group N=			No Exercise Group N=		
	Placebo (n=)	Low Range (n=)	High Range (n=)	Placebo (n=)	Low Range (n=)	High Range (n=)
Fasting plasma glucose - mg/dl						
Fasting plasma insulin - uU/ml						
Fasting serum lipids - mg/dl (std) Total cholesterol LDL HDL Triglycerides						
Liver function test – U/L (std) ALP AST ALT Bili T mg/dL						
PSA – ng/mL (std)						
RBC – 10 ¹² L (std)						
HCT - % (std)						
Blood pressure - mm Hg (std) Systolic Diastolic						
SHBG - nmol/L (std)						
FSH - mIU/mL (std)						
LH mIU/mL (std)						
IGF-1 – ng/ml (std)						
IGFBP-3 - ng/ml (std)						
Prolactin - ng/ml (std)						
TSH - uIU/ml (std)						
Leptin - ng/ml (std)						
Previous exercise						
Total Testosterone – ng/dL (std)						
Bioavailable Testosterone -						
History of T supplementation (no use 6 mo prior for eligibility)						
AUA						

Table 1c Baseline measures of physical function by randomization group

	Exercise Group N=			No Exercise Group N=		
	Placebo (n=)	Low Range (n=)	High Range (n=)	Placebo (n=)	Low Range (n=)	High Range (n=)
csPFP total						
csPFP subcategories: Upper body strength Upper body flexibility Lower body strength Balance and coordination Endurance						
Stair Climbing						
6 min walk						

Table 1d Baseline measures of strength by randomization group

	Exercise Group N=			No Exercise Group N=		
	Placebo (n=)	Low Range (n=)	High Range (n=)	Placebo (n=)	Low Range (n=)	High Range (n=)
1RM	92.3 (15.5)					
Power Rig						
Peak Torque (biodex)**						
Knee (slow)						
Knee (fast)						
Ankle (slow)						
Ankle (fast)						
Shoulder (slow)						
Shoulder (fast)						
Elbow (slow)						
Elbow (fast)						

** Question: how many angles?

Table 1e Baseline measures of body composition by randomization group

	Exercise Group N=			No Exercise Group N=		
	Placebo (n=)	Low Range (n=)	High Range (n=)	Placebo (n=)	Low Range (n=)	High Range (n=)
Weight - kg (std) (from DXA)						
Body Mass Index – kg/m ² (std)**						
Body Mass Index **	<20 20-24 25-29 30-34 35+	%tile <5 th 5-85 th 85-95 th >95 th				
Body circumferences – cm (std) Waist Hip Arm Thigh						
Body composition % Lean mass % Fat mass, etc						
BMD (DXA): Spine Total Hip Trochanter Femoral Neck Shaft						
CT: Abdominal – 1 Abdominal – 2 Thigh						

Table 1f Baseline questionnaire responses by randomization group

	Exercise Group N=			No Exercise Group N=		
	Placebo (n=)	Low Range (n=)	High Range (n=)	Placebo (n=)	Low Range (n=)	High Range (n=)
SF-36						
AUA						
Sexual performance						
Mood						
Depression						
Yale PAQ						
MACKL?						
Cog function?						

Draft document 12/13/2010

Tables 2 and 3: Attached as separate (pdf) document