STURDY

STUDY TO UNDERSTAND FALL REDUCTION AND VITAMIN D IN YOU

STATISTICAL ANALYSIS PLAN

VERSION 1.2

26 MAY 2017

NCT02166333

Table of Contents

Document history

Version 1.0 (29Oct2015)
Original version (no history)

Version 1.1 (15Mar2016)
Corrected terminology in section 3.1.1 (changed ‘median’ time to event to ‘mean’ time to event).

Version 1.2 (26May2017)
Corrected burn-in definition (from 100th participant randomized to 100th participant randomized to non control group) and adjustment interval (when each 100th person randomized to a non control group has completed 6 months of follow-up), updated serum vitamin D eligible range, added more details on number of simulations.

1 Introduction

STURDY is a seamless, two-stage, Bayesian response-adaptive, randomized trial to select the best dose of vitamin D for the prevention of falls in 1200 community-dwelling adults, ages 70+, who are at increased risk of falling and who have a serum 25-hydroxyvitamin D (25(OH)D) level of 10-29 ng/ml. In Stage 1 of the trial, participants will be randomly assigned to one of four vitamin D3 (cholecalciferol) doses: 200 IU/d (comparator), 1000 IU/d, 2000 IU/d, or 4000 IU/d, with assignment probabilities that will vary as falls are reported and entered into the database. This design will select the best dose of vitamin D for prevention of falls, or confirm the futility of distinguishing any differences among the doses for fall prevention. If a best dose is selected, participants from Stage 1 of the trial in the comparator (200 IU/d) and best dose groups will remain at their assigned doses, and participants in the other two dose groups will be dose-adjusted to the best dose. Subsequent participants enrolled during Stage 2 of the trial will be randomized to either the comparator or best dose group. In both Stages 1 and 2, all participants will take study pills and be followed for outcomes for two years or until the end of the study, whichever comes first. In-person follow-up visits will occur at 3, 12, and 24 months post-randomization.

The **primary outcome** is time to (a) first fall (defined as any fall, slip, or trip in which the participant loses his or her balance and lands on the floor or ground or at a lower level) or (b) death, whichever comes first. The **secondary outcome** is change in gait speed. The specific aims are listed below and in the STURDY protocol.

1.1 Primary aims

1. Use a Bayesian adaptive randomized clinical trial design to select the best of 4 doses of vitamin D supplements (200, 1000, 2000, or 4000 IU/d) for prevention of falls or to confirm the futility of distinguishing any differences among the doses for fall prevention (Stage 1).
2. If a best dose is selected, use a parallel group randomized clinical trial to obtain preliminary estimates of the efficacy of the best dose compared to the 200 IU/d dose for prevention of falls (Stage 2).

1.2 Other aims

3. Estimate the effects of vitamin D supplements on the outcomes of:
    a. gait speed
    b. 25(OH)D blood level
    c. physical performance measures (e.g., balance, grip strength, and SPPB score)
    d. physical activity level based on accelerometry
    e. other fall outcomes (e.g., multiple falls and fall rates, indoor falls, outdoor falls, injurious falls, falls that result in fractures, and falls that prompt emergency medical care)
    f. other non-fall outcomes (e.g., blood pressure, orthostatic blood pressure, onset of frailty, health-related quality of life)
4. Estimate the shape of the dose-response relationship of vitamin D supplements with the risk of falls and other outcomes
5. Estimate the shape of the dose-response relationship using other measures of vitamin D status with risk of falls and other outcomes:
    a. achieved 25(OH)D blood level at 3 months
    b. total dietary intake of vitamin D (food, usual supplements, and randomized dose)
6. Explore potential differences in outcomes by subgroups:

   a. key subgroups that may have increased benefit from vitamin D supplementation: blacks, those with baseline 25(OH)D of 10-19 ng/ml, and those with objective evidence of low physical function
   b. other subgrouping variables of interest: baseline supplement use, baseline vitamin D intake, gender, age, body mass index (BMI), medications, and frailty status
7. Explore the extent to which the relationship between vitamin D supplementation and falls is mediated by:
   a. physical activity levels assessed using accelerometry
   b. physical performance measures
8. Quantify safety concerns related to vitamin D supplementation by examining the frequency of hypercalcemia and acute nephrolithiasis.
9. Establish a well-characterized cohort of individuals at high risk of falls, with an associated biorepository, from which further research questions and creative ancillary studies can be conducted.

2 Data Source

Data will be entered by clinic staff into a web-based data system and the Data Coordinating Center (DCC) will oversee data storage. Backup files of the database will be generated and stored at regular intervals in a secure, off-site location, to permit regeneration of the database in the event that it is destroyed. Freeze dates for data sets created for interim and publication analyses will be documented.

3 Primary outcome: Time to first fall or death – establishing a best dose and testing for superiority

3.1 Randomization and primary analysis

3.1.1 Stage 1-- Dose Finding

Following methods outlined by Berry, et al,[1] we will employ a Bayesian adaptive design in Stage 1. The Bayesian adaptive design for Stage 1 will assign participants to dose 200 IU with a fixed 50% probability and doses 1000 IU, 2000 IU, or 4000 IU with varying probability using a response adaptive randomization scheme. Stage 1 will start with a burn-in period. The burn-in period is defined as the period from randomization of the 1st participant to the time that the 100th participant randomized to a non control group has been followed for 6 months. The length of the burn-in and the number of participants randomized during the burn-in will depend on recruitment. Assuming 25-30 participants are randomized each month, the burn-in period is estimated to be ~14 months (8 months to randomize 100th participant to a non control group, followed by 6 months of follow-up). We estimate the number of participants randomized during the burn-in to be about 300. At the end of the burn-in period, the dose group assignment probabilities will be adjusted based on the primary outcome of first fall or death, and thereafter, adjusted after every 100th participant randomized to any non-control group has completed 6 months of follow-up using the cumulative data available on first fall or death outcomes (ie, after the 200th, 300th, 400th, etc participant randomized to a non control group has completed 6 months of follow-up).

The assignment probabilities will be calculated using Bayes Theorem, given the prior distribution of time to fall or death and the fall data observed as of the time of the probability updates. Following common practice and implementation of the ARand software[2] from Berry et al,[1] we assume inverse

gamma distributions[3] for the mean times to fall for both the posterior and conjugate prior distributions and we specify a maximum sample size of 1200 participants for the Stage 1 dose finding. We assume the event time is exponentially distributed with mean λ, and that λ is distributed as an inverse Gamma with shape parameter α and scale parameter β^{l}. The prior for α can be thought of as the prior number of events and the prior for β can be thought of as the prior exposure time. Based on an informal synthesis of the previous literature,[4-15] we expect the 6 month fall proportion to be approximately 20%. Assuming exponentially distributed event time, the survival function at time x is exp (λx) with a mean survival time of $1 / \lambda$.[16] So a 6 month fall proportion of 20% corresponds to a hazard $\lambda = -ln(1-0.2) / 6 = 0.037$ and a mean time to event of 1 / 0.037 = 27 months. Therefore we set the prior of α to 2 in each dose group and the prior of β to 27 for each dose. This is a vaguely informative prior distribution chosen so that the mean time to first fall ($\beta / [\alpha - 1]$) is 27 months and corresponds to a 6 month fall proportion of 20%.

The posterior distribution for λ at a given time t is an inverse Gamma with shape α + number of events observed up to t and scale β + total exposure time up to time t. The "best" dose x among the three doses $x, y, z,$ i.e., 1000, 2000 and 4000, is the dose with the longest time to first fall (i.e., $P(\lambda_x > \max(\lambda_y, \lambda_z))$.[1;3] The predictive probability $P(\lambda_x > \max(\lambda_y, \lambda_z))$ for doses 1000, 2000 and 4000 will be estimated using a Monte Carlo algorithm with 200,000 simulations and pre-defined random number seeds. The number of simulations is based on the desire to have substantial precision of the estimate of the predictive probability. If a predicted probability of 0.025 or less results in an early loser, the exact 95% binomial confidence interval for 5000 out of 200,000 is 0.0243% to 0.0256%, yielding a precision of 0.0014.

For the next group of treatment assignments, the probability of being assigned to the comparator (200 IU) will always be 0.50 and the probability of assignment to the remaining doses 1000, 2000, and 4000 IU will be set to the predictive probability (of being the "best" dose) using the data collected up to time t, multiplied by 0.50. Again following advice in Berry et al., we also specified the following decision rules leading to finding a best dose (early winner and early losers) and proceeding to Stage 2 before follow-up is complete on 1200 participants:

- *Early loser*: If the Bayesian probability that the 1000, 2000 or 4000 IU dose is the best of those 3 doses for fall prevention is less than 2.5%, then accrual to that dose group will be suspended. The suspended dose can return to active status if, after the next update, the early loser rule is no longer true for the suspended dose.
- *Early winner*: If the Bayesian probability that the 1000, 2000 or 4000 IU dose is the best of those 3 doses for fall prevention is greater than 95%, then Stage 1 will stop early with that dose declared as the best and that dose will proceed to Stage 2.

We expect to find a best dose and terminate Stage 1 using fewer than 1200 participants under a variety of assumed dose-response scenarios (see Tables 1 and 2) and therefore we expect to proceed to Stage 2 (efficacy). However, we recognize that it is possible that the dose finding stage continues until the end of the STURDY trial with no Stage 2 data collected.

3.1.2 Stage 2 – Efficacy

If an early winner is declared, then both the comparator (200 IU/day) and "best" dose group will continue seamlessly into Stage 2 with continued recruitment and continued follow-up for all primary and secondary outcomes. The remaining two non-"best" dose groups will be dose-adjusted (up or down) to the "best" dose with all falls and all accumulated time counted in the primary analysis (intention to treat). These dose-adjustment groups will also provide important secondary information on the benefits, if any, of dose-adjustment.

Our null hypothesis that there is no effect of treatment for the three doses > 200 IU versus comparator (200 IU) will be tested by pooling all participants in the > 200 IU doses and comparing the pooled data with those in the comparator dose using the logrank test, based on all participant data at completion of the study. This hypothesis does not depend on selecting a best dose and continuing to Stage 2 and *will be performed even if a winner is not declared in Stage 1*. If this hypothesis is rejected, further preplanned tests will be conducted to compare individual dose levels, and to test for trends.

3.2 Power and sample size

3.2.1 Stage 1-- Dose Finding

We performed simulated trial scenarios using the stopping rules and assignment rules described above with varying dose-response assumptions (flat, decreasing, increasing, U-shaped). The Bayesian adaptive design assumed a maximum N=1200, accrual of 25 participants per month over 48 months. The number of months needed to arrive at a decision and end Stage 1 and various assumptions about the shape are shown below. We are allowing an approximately 20% probability of selecting a dose in the flat scenario below (similar to type I error) during the selection stage because we are controlling type I error at 5% in the final test of superiority when all follow-up is complete. A type I error rate of 20% is higher than the usual type I error allowed in phase 3 studies but lower than the type I error allowed in many phase 2 cancer selection trials designed to pick from a set of possible treatments to go on the phase 3 testing.[17-19]

Table 1 - Simulations for power and type I error

Scenario	Shape of dose response	Assumed proportion with fall at 6 months	Median # of participants to select dose in Stage 1 (Q1, Q3)	Probability of selecting true best dose	Probability of selecting the wrong dose
1	Decreasing linear 1	20% in 200 IU 15% in 1000 IU 11% in 2000 IU 8% in 4000 IU	800 (575, 1125)	82%	2%
2	Flat	20% in 200 IU 20% in 1000 IU 20% in 2000 IU 20% in 4000 IU	1200 (1200, 1200)	n/a	17%

Figure 1a - Distribution of number of participants in Stage 1 for scenarios 1 and 2

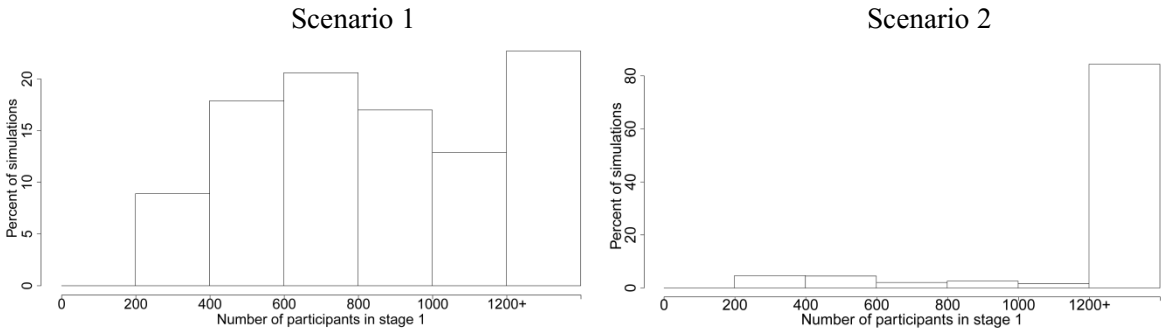

Table 2 - Simulations for other possible scenarios

Scenario	Shape of dose response	Assumed proportion with fall at 6 months	Median # of participants to select dose in stage 1 (Q1, Q3)	Probability of selecting true best dose	Probability of selecting the wrong dose
3	Decreasing linear 2	20% in 200 IU 17% in 1000 IU 14% in 2000 IU 11% in 4000 IU	925 (575, 1200)	70%	2%
4	Decreasing linear 3 (15% fall in control)	15% in 200 IU 12% in 1000 IU 9% in 2000 IU 6% in 4000 IU	775 (500, 1075)	87%	2%
5	Decreasing linear 4 (25% fall in control)	25% in 200 IU 22% in 1000 IU 18% in 2000 IU 15% in 4000 IU	1000 (600, 1200)	61%	4%
6	U-shaped, 2000 best	20% in 200 IU 19% in 1000 IU 14% in 2000 IU 19% in 4000 IU	875 (575, 1200)	80%	2%

3.2.2 Stage 2 – Efficacy

As discussed above, the sample size needed for Stage 1 will be variable. After Stage 1 is complete, newly enrolled participants will be randomized in a 1:1 ratio to either the "best" dose or the comparator dose (200 IU). New participants will be enrolled until a total of 1200 participants (Stage 1 and Stage 2) have been randomized. Existing participants in the 200 IU (comparator) and the best group will continue in their assigned group. Existing participants in the other dose groups will be re-assigned to the best group for Stage 2 follow-up. The primary outcome variable is time to first fall or death in the comparator (200 IU) group compared to time to first fall or death in the best dose group (data from the group identified as best in Stage 1 will be combined with data from the participants randomized to that dose in Stage 2 and with data from the participants originally assigned to the two non-best doses who are dose adjusted up or down to the best dose in Stage 2). A two-sided logrank test with 1200 subjects (600 controls vs. 600 in the best dose group) achieves 92% power, assuming Type I error=0.05, hazard ratio=0.75, 10% dropouts, and assuming participants enrolled in the first 2 years are closed out after 2 years (varying follow-up from 6 months to 2 years).

3.3 Sensitivity analyses

3.3.1 Stratification by time period

There is potential for bias in response adaptive randomization (RAR) due to a drift over the time course of the trial in the risk of falls combined with changes in randomization probabilities due to the response adaptive design. Chappell (2007),[20] in a National Cancer Policy Forum Workshop, gives a brief, but useful exposition of this potential bias when response adaptive randomization is used and there is a drift in the outcome rates. He ends with two conclusions:

1. "Results from response adaptive randomized phase II studies may be superior to those produced from nonrandomized designs."
2. "Due to fears of bias from confounding with time trends, RAR cannot substitute for true randomization in confirmatory trials."

We believe that the risk of this bias is low in the STURDY trial, since there is no current evidence that the risk of falling in the elderly is changing rapidly enough to affect the risk of falls over the 2-5 year dose-finding period of adaptive randomizations. Nonetheless, we will monitor data on potential drifts with elapsed time in the risk of falls in the comparator group (approximately half the participants at any given calendar time) during both Stage 1 and 2. In addition, we will monitor and report to the Data and Safety Monitoring Board how the event rates and treatment effect estimates are changing by time period defined by the group of participants enrolled with the same randomization probabilities (approximately every 200 participants after burn-in). As a sensitivity analysis, we will construct estimates of the treatment effects stratified by the time period. Karrison, et al[21] demonstrated that the stratified estimate maintains the type I error rate under a variety of allocation rules.

After the best dose is selected in Stage 1 (assuming finding a best dose is not futile), the second, efficacy stage of our trial will compare approximately equal numbers of participants assigned to the comparator group versus participants who received the best dose for all or part of the trial follow-up, which avoids the potential bias for the final comparison.

3.3.2 Limiting to "best" dose in both stages

The primary outcome variable of time to first fall or death in the comparator (200 IU) group will also be compared to time to first fall or death using only participants assigned to the dose group selected as the "best" dose in Stage 1 using a two-sided logrank test. The number of participants in this comparison will vary, depending on the number assigned to the "best" group during Stage 1.

3.3.3 Adjustment for fall risk factors

Since adjusting for baseline covariates known to be associated with the outcome might increase the efficiency of the estimate of the treatment effect,[22;23] we will do sensitivity analyses adjusting for the following baseline covariates: age, history of falls in the last year, gait speed and the use of a walking aide device, that are known to increase the risk of falls.[24] We will use Cox proportional hazard regression to estimate the hazard ratio of the comparator (200 IU) versus the combined (>200 IU) vitamin D dose groups controlling for the baseline covariates. Other potential confounders that we will monitor for baseline imbalance are gender, race, baseline BMI and baseline physical performance (SPPB score).

3.3.4 Time to first fall censoring at death and time to death

The composite primary outcome, time to first fall or death, consists of two components, fall and death, and the death outcome will prevent occurrence of the fall outcome (competing risk). We will compare the time to first fall with censoring at time of death for the comparator (200 IU) versus the combined (> 200 IU) groups using a logrank test. We will also compare the risk of death in the comparator versus combined groups.

3.3.5 Achieved vitamin D levels

We will calculate the achieved vitamin D levels using the participants' measurements of 25(OH)D blood level at 3 months. We will estimate the effect of the continuous measure of achieved vitamin D on the incidence of falls or death using Cox proportional hazard regression. We will test the null hypothesis that there is no relationship between achieved vitamin D and incidence of falls or death.

4 Secondary outcome: Change in gait speed

4.1 Analysis

We will calculate the change in gait speed measured by SPPB from baseline to month 3 for each of the 1200 participants. We will test the null hypothesis that there is no difference in the change in gait speed in the comparator (200 IU) versus the combined (>200 IU) groups using a t test or Wilcoxon rank sum (if the t test model assumptions are not met). If this hypothesis is rejected, further preplanned tests will be conducted to compare individual dose groups (4000, 2000, 1000 versus 200) using a Bonferroni correction for p-values, and to test for trends. We will also inspect the difference between the target date for study procedures and the actual dates. We will conduct sensitivity analyses to exclude participants with actual follow-up visit dates that are more than three months from the target date.

In addition we will use longitudinal models to compare the gait speed at months 12 and 24 for the *subset* of participants enrolled in the first 36 months and 24 months, respectively (i.e., those who are expected to have at least 12 and 24 months of follow-up by design) using mixed effects models including baseline, months 3, 12 and 24 (when applicable) gait speed measures.

4.2 Power and sample size

With N=1200, the trial will have 90% power for detecting mean differences as small 0.2 SDs in outcomes such as gait speed between the comparator versus combined groups. Given that usual gait speed over 6 meters and 2.5 minutes of customary paced ("walk at your usual comfortable pace") averaged 1.1 (SD=0.2) m/s in sample of 420 older adults[25], our sample size is sufficient to detect minimum clinically important differences (MCIDs) as small as 0.04 m/s, which compares favorably with meaningful change and responsiveness in the 0.05-0.10 m/s range.[26]

5 Other secondary physical activity and performance outcomes

We will use the analysis strategy described in 4.1 for comparison of change in other continuous outcome measures such as grip strength, SPPB score, physical activity measured by accelerometry, systolic or diastolic blood pressure, orthostatic systolic or diastolic blood pressure, and health-related quality of life.

Improvement in balance will be defined as a binary variable with the event being a one or more point improvement in balance as measured by the SPPB (range: 0-4) compared to baseline. We will test the null hypothesis that there is no difference in the proportion of participants with improvement in balance in the comparator (200 IU) versus the combined (>200 IU) groups using a χ^2 test at months 3, 12 and 24. As described previously, if the overall hypothesis is rejected, further preplanned tests will be conducted to compare individual dose groups (4000, 2000, 1000 versus 200).

6 Other fall outcomes

6.1 Shape of the dose-response relationship of vitamin D supplementation on falls

We will use a 4-parameter logistic model for modeling dose-response of vitamin D supplementation versus the outcome of proportion of falls or death at 6 months. The 4-parameter logistic is derived from the classic 2-parameter logistic, adding extra parameters to allow the function to extend beyond 0 and 1. The parameters are estimated using nonlinear regression methods and the resulting functions fit a wide variety of dose-response shapes. The 4-parameter logistic for the expected response Y, given a dose X is $E(Y|X) = \theta_3 + \frac{\theta_4 - \theta_3}{1+\left(X/\theta_1\right)^{\theta_2}}$.[27;28] In this model, θ_3 and θ_4 are the maximum and minimum response levels. The concentration where the response is halfway between the maximum and minimum is sometimes called ED50 and shown as θ_1. The slope parameter measures the speed with which the curve rises between the minimum and maximum and is shown here as θ_2. We will calculate the asymptotic 95% confidence intervals for θ_2 to test the null hypothesis that the slope of the dose-response relationship is zero.

We will construct similar dose-response models for the relationship between achieved 25(OH)D and proportion of falls at 6 months as well as vitamin D dose and achieved 25(OH)D with other outcomes such as gait speed.

6.2 Injurious falls

We will use logrank methods, similar to those described above for the primary outcome, to compare the incidence of the outcomes of injurious falls, falls that result in fractures, and falls that prompt emergency medical care in the comparator versus the combined vitamin D dose groups and the individual vitamin D dose groups. Participants with falls that do not meet the outcome definition (non-injurious, no fracture, etc) will continue in the 'at risk' group for the fall outcome of interest.

6.3 Fall rates

We will continue to collect data on falls after a participant experiences a first fall. Rates of falls (average number of falls per person per month of follow-up) will be computed for all falls, indoor falls, and outdoor falls including data for all 1200 participants. Negative binomial regression models will be used to compare the rates of falls in the comparator versus the combined vitamin D dose groups and for the comparator versus the individual vitamin D dose groups. We will compute unadjusted risk ratios of fall rates as well as adjusted risk ratios controlling for the covariates known to be associated with risk of fall listed above in section 3.3.3.

7 Potential effect modification

We will explore the extent to which the effect of vitamin D supplementation on the outcomes listed above differs in several subgroups of interest. We will compare the treatment effects in subgroups by testing for treatment (dose group) by subgroup interactions in the models described above. The pre-specified key subgroups of interest are blacks versus other races, those with baseline 25(OH)D of 10-19 ng/ml versus higher than 19 ng/ml, and those with objective evidence of low versus higher physical function at baseline (< 10 vs ≥ 10 SPPB total score). Other subgroups of interest are baseline supplement use (none vs. any), total baseline vitamin D intake (< 800 IU/d vs. ≥ 800 IU/d)[29], gender, age (< 80 years vs. ≥ 80 years), body mass index (BMI; < 18.5 vs. ≥18.5 to < 25 vs. ≥25 to < 30 vs. ≥ 30), medications classes (any antihypertensive [including diuretics, ACE inhibitors, beta blockers or calcium channel blockers] vs. no antihypertensive and diabetic drugs [insulin or oral anti-diabetic drugs] vs. no diabetic drugs), frailty status, and prior fallers versus no prior fall.

8 Mediation models

Mediation analysis can be used to identify the underlying mechanisms of interventions. Changes in the time-dependent covariates, physical activity as measured by accelerometry and physical performance measures, will be subsequently added to the Cox survival model to test our hypotheses that these changes would mediate or explain a significant portion of the vitamin D effect on falls.

To demonstrate statistical mediation according to accepted guidelines,[30;31] the vitamin D supplementation must lead to significant change in the potential mediator (physical activity or physical performance), and this supplement induced change must also explain a significant proportion of variance in the primary outcome, time to first fall. We will determine that mediator's total mediation effect using methods for survival outcomes described in Roth, 2012.[32] The proportion of the vitamin D effect that can be attributed to the mediator is the difference in magnitude between the vitamin D effect in the baseline covariate-adjusted model and the vitamin D effect in the mediation model that also includes the change score for that time-dependent predictor.

The total mediation effect for each mediator includes both the unique influence of that mediator and the effect it shares with other mediators, since changes in these mediators caused by the intervention are not necessarily independent of each other. Therefore, additional multivariate mediation models will be estimated in which multiple time-dependent changes are entered simultaneously. These models will allow us to quantify the proportion of the total vitamin D effect on falls that could be attributed to changes in the potential mediators collectively.

9 Missing data

We will employ recommended strategies to prevent missing data, based on published research (National Research Council 2010;[33] Mills et al, 2006;[34] Ross et al, 1999;[35] Booker et al, 2011[36]) and previous experience of the Data Coordinating Center:

- We have developed a data collection schedule that will minimize participant burden and we will follow participants according to the data collection schedule regardless of compliance with the study intervention;
- We will maintain frequent contact with the participants through visit reminder calls or notes;
- We will provide a 24-hour phone number that participants can contact for questions and support;
- We will create contiguous windows of time during which specific follow-up visits are allowed;
- We will employ rigorous training of clinic staff emphasizing the importance of:
    - Congenial interpersonal relationships between the participants and study staff ;
    - Using the consent process to ensure that potential participants understand the commitment that they are making;
    - Addressing concerns if participants are dissatisfied so that the participant will remain in the trial;
    - Collecting data even if a participant discontinues the study treatment or cannot come to the clinic because of other obligations.

We will collect data on reasons for study dropout. If participants who drop out of the study appear to be doing so for reasons related to the study, i.e., not missing completely at random, we will perform sensitivity analyses using methods that have been described,[37] such as multiple imputation techniques, best- and worst-case scenarios, and correlates of the drop out event included in the models. There are no standard statistical techniques for dealing with data that are missing not at random (MNAR). We will explore one or more of the sensitivity analyses for MNAR given in the NRC report.[33] Baseline characteristics of participants with missing measures will be compared between treatment groups.

10 Safety outcomes

We will compare the rate of adverse events by dose group using negative binomial models. The following events are adverse events: new disease diagnoses; worsening of pre-existing disease; patient symptom reports; falls that require medical treatment or evaluation. We will combine the reported AEs into Medical Dictionary for Regulatory Activities (MedDRA) system organ classes to look for increased risk of events by organ system. If differences by organ system are found, we will compare events by dose group for the MedDRA High Level Terms or Preferred Terms.

We will also compare the rates of all serious adverse events (defined according to the National Institute on Aging definition[38]) using negative binomial models. The following events will be considered serious adverse events: hospitalizations or prolongations of hospitalizations; deaths; kidney, bladder or ureteral stones; hypercalcemia defined as serum calcium $\geq$ 11.0 mg/dl; serum calcium $\leq$ 8.4 mg/dL; vitamin D $\geq$ 150 ng/mL; vitamin D < 10 ng/mL.

11 Interim monitoring

The NIA has appointed a multidisciplinary DSMB that will be responsible for the protection of the safety of participants enrolled in the trial. Details on the DSMB responsibilities, meetings and reports can be found in the DSMB charter.

At each interim analysis, the DSMB will make recommendations to the NIA about continuing, modifying, or stopping the trial. The DSMB will meet periodically to review interim reports and analyses derived from the accumulating data or related findings from sources external to the trial that may be needed to make recommendations to the NIA regarding: 1) overall efficacy and benefit/risk ratio, 2) efficacy and benefit/risk ratios within defined subsets of participants, and 3) overall and clinic-specific performance and data quality.

The DCC will prepare a report for each DSMB meeting. The report will have two parts:

Part 1 of the DSMB report: accumulating information relating to recruitment, data quality (e.g., data return rates, audit results), treatment compliance and baseline characteristics. Details on adverse events and harms based on pooled data may be presented.

Part 2 of the DSMB report: Efficacy and safety data, by dose group. Dose groups will be presented as blinded groups, i.e., dose A, dose B, etc.

There will be no formal interim analysis of the primary efficacy outcome. However, the DSMB will meet to review the proposed Bayesian design recommendations for probability changes after the burn-in period and before the changes to randomization probabilities are implemented. At this meeting, the DSMB will also review the fall rates by prior fall status and by groups based on the other fall screening questions. The crude fall rates will be presented at each DSMB meeting without hypothesis testing. If we hit a stopping rule in Stage 1, then the DCC will convene a meeting of the DSMB as soon as possible.

Literature Cited

1. Berry SM, Carlin BP, Lee JJ, Muller P. Bayesian Adaptive Methods for Clinical Trials. Boca Raton, FL: Taylor and Francis Group, LLC; 2011.

2. ARandDesktop [computer program]. Version 4.1.1. Houston, TX: 2015.

3. Cook JD. Numerical evaluation of gamma inequalities. 2006. Report No.: Technical Report UTMDABTR-001-06.

4. Bischoff-Ferrari HA, Dawson-Hughes B, Staehelin HB, et al. Fall prevention with supplemental and active forms of vitamin D: a meta-analysis of randomised controlled trials. BMJ 2009;339:b3692.

5. Bischoff-Ferrari HA, Dawson-Hughes B, Willett WC, et al. Effect of Vitamin D on falls: a meta-analysis. JAMA 2004;291:1999-2006.

6. Bischoff HA, Stahelin HB, Dick W, et al. Effects of vitamin D and calcium supplementation on falls: a randomized controlled trial. J Bone Miner Res 2003;18:343-51.

7. Broe KE, Chen TC, Weinberg J, et al. A higher dose of vitamin d reduces the risk of falls in nursing home residents: a randomized, multiple-dose study. J Am Geriatr Soc 2007;55:234-9.

8. Dukas L, Bischoff HA, Lindpaintner LS, et al. Alfacalcidol reduces the number of fallers in a community-dwelling elderly population with a minimum calcium intake of more than 500 mg daily. J Am Geriatr Soc 2004;52:230-6.

9. Bunout D, Barrera G, Leiva L, et al. Effects of vitamin D supplementation and exercise training on physical performance in Chilean vitamin D deficient elderly subjects. Exp Gerontol 2006;41:746-52.

10. Pfeifer M, Begerow B, Minne HW, et al. Effects of a long-term vitamin D and calcium supplementation on falls and parameters of muscle function in community-dwelling older individuals. Osteoporos Int 2009;20:315-22.

11. Pfeifer M, Begerow B, Minne HW, et al. Effects of a short-term vitamin D and calcium supplementation on body sway and secondary hyperparathyroidism in elderly women. J Bone Miner Res 2000;15:1113-8.

12. Zamboni M, Zoico E, Tosoni P, et al. Relation between vitamin D, physical performance, and disability in elderly persons. J Gerontol A Biol Sci Med Sci 2002;57:M7-11.

13. Flicker L, MacInnis RJ, Stein MS, et al. Should older people in residential care receive vitamin D to prevent falls? Results of a randomized trial. J Am Geriatr Soc 2005;53:1881-8.

14. Kalyani RR, Stein B, Valiyil R, et al. Vitamin D treatment for the prevention of falls in older adults: systematic review and meta-analysis. J Am Geriatr Soc 2010;58:1299-310.

15. Sanders KM, Stuart AL, Williamson EJ, et al. Annual high-dose oral vitamin D and falls and fractures in older women: a randomized controlled trial. JAMA 2010;303:1815-22.

16. Klein JP, Moeschberger ML. Survival analysis: techniques for censored and truncated data. New York, NY: 1997.

17. Ratain MJ, Sargent DJ. Optimising the design of phase II oncology trials: the importance of randomisation. Eur J Cancer 2009;45:275-80.

18. Korn EL, Freidlin B, Abrams JS, et al. Design issues in randomized phase II/III trials. J Clin Oncol 2012;30:667-71.

19. Rubinstein L, Crowley J, Ivy P, et al. Randomized phase II designs. Clin Cancer Res 2009;15:1883-90.

20. Chappell R. Comments on the Controversy over Response-Adaptive Randomization. National Cancer Policy Forum Workshop. Washington, DC 2007.

21. Karrison TG, Huo D, Chappell R. A group sequential, response-adaptive design for randomized clinical trials. Control Clin Trials 2003;24:506-22.

22. Lu X, Tsiatis AA. Improving the efficiency of the logrank test using auxiliary covariates. Biometrika 2008;95:679-94.

23. Zhang M, Tsiatis AA, Davidian M. Improving efficiency of inferences in randomized clinical trials using auxiliary covariates. Biometrika 2008;64:707-15.

24. Deandrea S, Lucenteforte E, Bravi F, et al. Risk factors for falls in community-dwelling older people: a systematic review and meta-analysis. Epidemiology 2010;21:658-68.

25. Schrack JA, Simonsick EM, Chaves PH, et al. The role of energetic cost in the age-related slowing of gait speed. J Am Geriatr Soc 2012;60:1811-6.

26. Perera S, Mody SH, Woodman RC, et al. Meaningful change and responsiveness in common physical performance measures in older adults. J Am Geriatr Soc 2006;54:743-9.

27. Dmitienko A, Chuang-Stein C, D'Agostino B. Pharmaceutical statistics using SAS: a practical guide. Cary, NC: SAS publishing; 2007.

28. Gottschalk P, Dunn J. The five-parameter logistic: A characterization and comparison with the four-parameter logistic. Analytical Biochemistry 2005;343:54-65.

29. IOM (Institute of Medicine). Dietary reference intakes for calcium and vitamin D. Committee to Review Dietary Reference Intakes for Calcium and Vitamin D. Washington, DC: National Academies Press, Institute of Medicine; 2011.

30. Shrout PE, Bolger N. Mediation in experimental and nonexperimental studies: new procedures and recommendations. Psychol Methods 2002;7:422-45.

31. Judd CM, Kenny DA, McClelland GH. Estimating and testing mediation and moderation in within-subjects designs. Psychol Methods 2001;6:115-34.

32. Roth DL, MacKinnon DP. Mediation analysis with longitudinal data. In: Newsom JT, Jones RN, Hofer SM, editors. Longitudinal data analysis: A practical guide for researchers in aging, health, and social sciences.New York, Ny: Routledge; 2012. p. 181-216.

33. National Research Council. The Prevention and Treatment of Missing Data in Clinical Trials. Panel on Handling Missing Data in Clinical Trials. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press; 2010.

34. Mills E, Wilson K, Rachlis B, et al. Barriers to participation in HIV drug trials: a systematic review. Lancet Infect Dis 2006;6:32-8.

35. Ross S, Grant A, Counsell C, et al. Barriers to participation in randomised controlled trials: a systematic review. J Clin Epidemiol 1999;52:1143-56.

36. Booker CL, Harding S, Benzeval M. A systematic review of the effect of retention methods in population-based cohort studies. BMC Public Health 2011;11:249.

37. Shih W. Problems in dealing with missing data and informative censoring in clinical trials. Curr Control Trials Cardiovasc Med 2002;3:4.

38. National Institute on Aging. NIA adverse event and serious adverse event guidelines. 2011. Report No.: Version 2.