Contemporary Clinical Trials Communications 14 (2019) 100333

Contents lists available at ScienceDirect

Contemporary Clinical Trials Communications

journal homepage: www.elsevier.com/locate/conctc

A randomized placebo-controlled trial of vitamin D supplementation for reduction of mortality and cancer: Statistical analysis plan for the D-Health Trial

Mary Waterhouse[a], Dallas R. English[b], Bruce K. Armstrong[c], Catherine Baxter[a], Briony Duarte Romero[a], Peter R. Ebeling[d], Gunter Hartel[a], Michael G. Kimlin[e], Donald S.A. McLeod[a], Rachel L. O'Connell[f], Jolieke C. van der Pols[g], Alison J. Venn[h], Penelope M. Webb[a], David C. Whiteman[a], Rachel E. Neale[a,i,*]

[a] Population Health Department, QIMR Berghofer Medical Research Institute, Brisbane, Australia
[b] Melbourne School of Population Health, University of Melbourne, Cancer Epidemiology and Intelligence Division, Cancer Council Victoria, Melbourne, Australia
[c] School of Public Health, University of Sydney, Sydney, Australia
[d] Department of Medicine, School of Clinical Sciences, Monash University, Melbourne, Australia
[e] Health Research Institute, University of Sunshine Coast, Sippy Downs, Australia
[f] NHMRC Clinical Trials Centre, University of Sydney, Sydney, Australia
[g] School of Exercise and Nutrition Sciences, Faculty of Health, Queensland University of Technology, Brisbane, Australia
[h] Menzies Institute for Medical Research, University of Tasmania, Hobart, Australia
[i] School of Public Health, The University of Queensland, Brisbane, Australia

ARTICLE INFO

Keywords:
Statistical analysis plan
Randomized controlled trial
Vitamin D
Mortality
Cancer

ABSTRACT

Background: Many observational studies have reported an association between vitamin D and non-skeletal health outcomes. The D-Health Trial was launched to determine if supplementing the older population with high monthly doses of Vitamin D can prevent cancer and premature mortality. The intervention is ongoing but here we provide a detailed statistical analysis plan for the primary and secondary outcomes of the D-Health Trial.
Methods/design: The D-Health Trial is a double-blind, randomized, placebo-controlled trial. Between February 2014 and May 2015, 21,315 people were randomized in a 1:1 ratio to receive monthly doses of either 60,000 IU of cholecalciferol (vitamin D_3) or placebo for five years. The primary outcome is all-cause mortality and the secondary outcomes are total cancer incidence and colorectal cancer incidence. These will be ascertained via linkage to death and cancer registries. The primary analysis for each outcome will follow an intention-to-treat approach; we will use flexible parametric survival models to investigate the association between supplementation and time to an event. We describe in detail sophisticated secondary analyses that consider non-compliance and contamination due to off-study supplementation.
Conclusions: Publication of this statistical analysis plan in advance of the intervention's completion, and adherence to it, will avoid data-driven analyses of the primary and secondary outcomes and ensure robust reporting of outcomes.
Clinical trial registration number: Australian New Zealand Clinical Trials Registry: ACTRN12613000743763. Registered on 4 July 2013.

1. Introduction

Many observational studies have reported inverse associations between serum 25-hydroxy vitamin D (25(OH)D) concentration (used as a measure of vitamin D status) and risk of non-skeletal health outcomes such as cancer, cardiovascular disease, diabetes and mortality [1,2]. However, findings from observational studies may be a result of reverse causality or uncontrolled confounding, so cannot be used to reliably infer that increasing 25(OH)D concentrations in the population (through supplementation or food fortification) would be beneficial.

In randomized controlled trials, vitamin D supplementation has not reduced the incidence of cancer [1,3–7]. The largest of these trials, the

* Corresponding author. QIMR Berghofer Medical Research Institute, Locked Bag 2000, Royal Brisbane Hospital, QLD, 4029, Australia.
E-mail address: rachel.neale@qimrberghofer.edu.au (R.E. Neale).

https://doi.org/10.1016/j.conctc.2019.100333
Received 29 October 2018; Received in revised form 25 January 2019; Accepted 6 February 2019
Available online 20 February 2019

Table 1
Tertiary outcomes of the D-Health Trial and the data sources used to derive them.

Outcome	All participants				Random sample of participants		
	Annual survey[a]	MBS database[b]	PBS database[b]	Hospital data[c]	Phone interview	Diary	Blood sample
Newly diagnosed health conditions[d]							
Anxiety	X		X				
Arthritis	X		X				
Depression	X		X				
Diabetes	X		X				
Hypercalcemia[e]	X						
Hypercholesterolemia	X		X				
Hyperparathyroidism	X		X	X			
Hyperthyroidism	X		X	X			
Hypothyroidism	X		X				
Kidney disease	X			X			
Osteoporosis[f]	X		X	X			
Parkinson's disease	X		X	X			
Psoriasis	X		X	X			
Sarcoidosis	X			X			
Glaucoma	X		X				
Insomnia	X		X				
Sleep apnea	X	X					
Non-alcoholic fatty liver disease	X			X			
Polymyalgia rheumatica	X						
Dementia/cognitive impairment/Alzheimer's disease	X		X	X			
Emphysema/chronic obstructive pulmonary disease/bronchiectasis	X		X	X			
Number of hospitalisations				X			
Acute events[g]							
Kidney stones	X			X			
Cataract removal	X			X			
Broken bones	X			X		X	
Joint repair or replacement	X			X			
Gall stones/Cholecystectomy	X			X			
Hysterectomy	X			X			
Cardiovascular outcomes							
Arrhythmia[d]	X		X	X			
Cerebrovascular disease[g,h]	X			X			
Coronary artery disease[i]	X		X	X			
Hypertension[d]	X		X				
Thrombosis[g]	X		X	X			
Quality of life/health and well-being (12-item Short Form Health Survey; 5-point Likert scale for overall health status, quality of life, memory, and teeth and gum health)	X						
Depressive symptoms (Patient Health Questionnaire- 9)	X						
Pain (Pain Impact Questionnaire-6)	X						
Sleep quality (Pittsburgh Sleep Quality Index)	X						
Acute respiratory tract infection[j]	X		X			X	
Infections	X		X	X			
Falls[k]	X					X	
Memory and cognition (Telephone Interview for Cognitive Status – modified)					X		
Urinary function (International Prostate Symptom Score (Men); Questionnaire for Urinary Incontinence Diagnosis (Women))	X						
Erectile dysfunction	X						
Treatment of keratinocyte cancer		X					
Telomere length							X

[a] Annual surveys are completed at the end of each year of the intervention.

[b] The Medicare Benefits Scheme (MBS) and Pharmaceutical Benefits Scheme (PBS) databases are managed by Medicare Australia. Records of consenting participants will be linked to these databases. The MBS database records health services provided outside the public hospital system. The PBS database records almost all medications prescribed outside a public hospital setting. Medications will be used as proxies for diagnosis. A first linkage captured data up to and including 31 December 2017. A second linkage is scheduled for 2021.

[c] Linkage to hospital admitted patient datasets.

[d] For conditions with specific medical and/or surgical treatments, PBS data regarding dispensing of drugs and/or hospital data regarding procedures will be used in preference to annual survey data to derive the timing of first diagnosis.

[e] We follow up diagnoses of hypercalcemia with medical records where possible.

[f] Hip and/or vertebral fracture(s) may be used to confirm a diagnosis of osteoporosis.

[g] Hospital data regarding procedures used to the treat the event will be given preference to annual survey data when deriving the number and timing of events.

[h] A person will be classified as having cerebrovascular disease if they have a transient ischaemic attack and/or a stroke.

[i] A person will be classified has having coronary artery disease (CAD) if they have symptoms of the disease (angina, myocardial infarction) and/or treatment for the disease (coronary stenting, coronary artery bypass graft, angioplasty) or if they are documented as having a diagnosis of CAD in the hospital data.

[j] Annual surveys ask whether or not a participant had a cold/runny nose/sore throat/the flu in the last month. Diaries use a scoring system of respiratory tract infection-related symptoms; these are completed daily over 8 weeks during winter.

[k] Annual surveys ask: whether or not a participant had a fall in the last month; and how many times a participant fell in the last 12 months. Diaries are used to indicate whether or not a participant fell during each day of a 3 month period; a fall is described as unintentionally coming to rest on the floor, the ground or other

lower level, regardless of what caused the fall. For each fall that is documented in the diary, participants also provide information on whether they sought medical help and injuries sustained as a result of the fall.

Women's Health Initiative trial, was designed to be adequately powered to detect an absolute difference in colorectal cancer incidence of 22% over 8 years [5]. The study did not observe a reduction in the incidence of invasive colorectal cancer (hazard ratio (HR), 1.08; 95% confidence interval (CI), 0.86 to 1.34). However, both the dose used (400 international units (IU)/day) and adherence were low, with only 60% of participants taking at least 80% of their study medication. Further, the study included only healthy, post-menopausal women, and their baseline intake of vitamin D from supplements was twice the national average. The large-scale Vitamin D and Omega-3 Trial (VITAL) found that supplementation with vitamin D at a dose of 2000 IU/day was not associated with lower incidence of invasive cancer of any type (HR, 0.96; 95% CI, 0.88 to 1.06) [7]. However, the study found that body mass index (BMI) may modify the effect of vitamin D, and amongst people with BMI $< 25\ kg/m^2$, incidence was lower in the vitamin D group than in the placebo group [7]. VITAL recruited men and women and the median follow-up was 5.3 years.

Meta- and pooled analyses of trials have found that vitamin D supplementation results in a small decrease in all-cause mortality [1,4,8], but the generalizability of these results is uncertain. Many studies were designed to assess falls and bone health, so included primarily women, and a substantial proportion of participants lived in aged care facilities. Mortality was usually a secondary outcome.

We launched the D-Health Trial [9] in response to calls for large, well-designed, population-based trials of vitamin D supplementation [8]. The D-Health trial aims to determine if monthly high-dose vitamin D supplementation of the general older population can prevent cancer and premature mortality, with results having the potential to influence public health policy.

For transparency of future reporting of results from the trial, we now provide a detailed statistical analysis plan for the primary (all-cause mortality) and secondary (total cancer incidence and colorectal cancer incidence) outcomes. Tertiary outcomes are also briefly discussed.

2. Methods

2.1. Brief trial overview

Detailed trial methods and baseline characteristics have been published [9]. Briefly, the D-Health Trial is a double-blind, randomized, placebo-controlled trial of monthly high-dose vitamin D supplementation in older Australian adults. Participants will be supplemented for 5 years, after which time we will continue to capture cancer and mortality outcomes through linkage with registries. The QIMR Berghofer Medical Research Institute Human Research Ethics Committee approved the trial and all participants gave written or online consent.

We invited 421,207 Australians aged 60–79 years to participate, using the Commonwealth electoral roll as a sampling frame, and also allowed volunteers to join the trial. We excluded people who had a history of osteomalacia, sarcoidosis, hyperparathyroidism, hypercalcemia, or kidney stones, or who were taking more than 500 IU of supplementary vitamin D per day. Between February 2014 and May 2015, the 21,315 people recruited (1,896 were volunteers) were randomized in a 1:1 ratio to receive monthly doses of either 60,000 IU of cholecalciferol (vitamin D_3) or placebo for five years. Randomization occurred within strata of age (60–64; 65–69; 70–74; 75 + years), sex, and state of residence (New South Wales, Queensland, South Australia, Tasmania, Victoria, Western Australia) at baseline.

2.2. Outcomes

The primary outcome is all-cause mortality. The secondary outcomes are total cancer incidence (excluding incident keratinocyte cancers of the skin, which are registered by only one of the Australian state-based cancer registries) and colorectal cancer incidence. These outcomes are ascertained via linkage to death and cancer registries. Preliminary linkage occurred in 2017/18. A final linkage is scheduled for 2020/21, with possible further linkages dependent on continued funding.

Table 1 lists the tertiary outcomes and the data sources used to derive them. These were selected from a review of published reports of associations between vitamin D and health outcomes. Some outcomes are ascertained from multiple sources, enabling triangulation of the data. Data regarding provision of health services and drugs dispensed outside public hospitals are available for participants who consented to linkage of their study records with those of the Medicare Benefits Scheme (MBS) (95%) and/or Pharmaceutical Benefits Scheme (PBS) databases (92%). Many tertiary outcomes are captured using annual surveys, which participants are encouraged to complete even if they have withdrawn from the trial. Some outcomes are collected for a random sample of participants only (Table 1). More detailed definitions of tertiary outcomes will be given in separate publications.

2.3. Sample size calculation

We based our sample size calculations on cumulative risk of death over a 10-year period under the assumption that the trial cohort experiences 0.8 of the event rate of the general Australian population. The trial was designed to detect, with 80% power and using a significance level of 0.05, a mortality HR of 0.88 (based on a logrank test) over a ten-year period from the start of the trial. We allowed for a drop-out rate of 20% and assumed that, on average, participants would become non-adherent at the midpoint of the trial.

An HR of 0.88 corresponds to an absolute risk reduction of 1.5% or the equivalent of needing to treat 66 patients to prevent one death. It equates to an extra 3.1 years of life for 60 year olds and 1.2 years for 79 year olds. Currently, we are funded to ascertain primary and secondary outcomes only over a six-to seven-year period, corresponding to a second data linkage occurring in 2020/21.

2.4. Compliance with study medication and use of off-trial vitamin D supplements

Although we restricted enrolment to people taking 500 IU or less of supplementary vitamin D per day, once enrolled we allow participants to remain in the trial provided that they take no more than 2000 IU/day of off-study vitamin D supplementation. This minimizes missing participant-reported information about outcomes and supplement use while ensuring participants take no more than the equivalent of 4000 IU per day, recommended as the tolerable upper intake level by the United States Institute of Medicine [10]. To estimate compliance and use of additional vitamin D supplements, the annual survey asks participants to report the number of study capsules taken during the previous 12 months (none, 1–3, 4–6, 7–9, 10–12), and intake of off-study vitamin D supplementation. We also ask participants to contact us to report any changes to off-study supplementation. Additionally, each year (beginning one year after baseline) we measure 25(OH)D concentration in a random sample of approximately 350 people from each trial arm.

2.5. Statistical analyses

2.5.1. Primary and secondary outcomes

2.5.1.1. General principles. A time-to-event approach will be used with time since randomization as the time axis. Follow-up will begin at date of randomization and end at the date of the event of interest, death

(when not the event of interest), final linkage, or loss to follow-up (e.g. for people who emigrate from Australia during follow-up), whichever comes first. For the primary outcome, the event is death from any cause. For the secondary outcomes, the events are: (1) first diagnosis of any cancer (excluding keratinocyte cancer); and (2) first diagnosis of colorectal cancer. We will fit separate models for each outcome. When analyzing a secondary outcome, if a person dies without having had a cancer diagnosis, then their data will be treated as censored at the time of death.

2.5.1.2. Primary analysis. The primary analysis will follow an intention-to-treat approach, using data from all people who were randomized, excluding the few (currently N = 5) participants who have withdrawn and requested destruction of their data. Participant flow will be shown using a CONSORT diagram [11]. We will report the number and percentage of events within each randomization group.

We will use a flexible parametric survival model (FPSM) [12] to estimate the association between an outcome and randomization group. We will model the baseline log cumulative hazard function using a restricted cubic spline with two internal knots (placed at the 33rd and 67th percentiles of the uncensored log survival times). To allow the estimated HR to vary with time, we will include an interaction between randomization group and a restricted cubic spline with one internal knot (placed at the median of uncensored log survival times). The choice of spline functions follows the recommendations made by Royston and Parmar [12,13]. The model will include the randomization strata of age, sex, and state of residence at baseline as non-time-dependent covariates.

We will plot the estimated HR and its 95% CI as a function of time since randomization, and report the estimated HR (95% CI) at 2, 4 and 6 years post-randomization. We will also use the FPSM to produce plots (with 95% CIs) of estimated survival functions by randomization group, and difference in survival functions.

2.5.1.3. Interactions and subgroup analyses. For each outcome, we will investigate whether the effect of supplementation is modified by: age at baseline (<70 years, $\geq$ 70 years); sex; predicted baseline deseasonalized 25(OH)D concentration (using the following range of cut points: 30 nmol/l; 40 nmol/l; 50 nmol/l; 75 nmol/l; and the median); and BMI ($<25\,kg/m^2$, $\geq 25\,kg/m^2$). For each factor we will report the P value from a likelihood ratio test comparing models with and without an interaction term (between randomization group and the factor of interest). Plots of the estimated HR (and 95% CI) for the association between supplementation and the outcome will be produced for each stratum of a factor.

2.5.1.4. Secondary analyses assessing the impact of compliance and off-study supplementation. We will perform secondary analyses that take into account the fact that some participants may not take all their study tablets, may consume off-study supplementary vitamin D, or may do both. Some of these analyses require data about intake of supplementary vitamin D in the period after we no longer have questionnaire data on this variable (e.g., in the post-trial passive/linkage period, and from the time of withdrawal for people who withdraw and don't continue to complete annual surveys). For simplicity, we will assume that from the date of last reported intake until the event or censoring, a person's intake remains constant at the dose last reported. Note that for these analyses, 'total intake' refers to intake from study tablets and off-study supplementation and does not include dietary intake. We will use the following approaches:

2.5.1.4.1. Per-protocol analysis. We will estimate the effect that would have been observed had all participants adhered to the protocol. We will define adherence as taking $\geq$80% of the study tablets and not taking more than 500 IU/day of off-study vitamin D supplementation at any time. We will 'artificially' censor participants at the time that they first report being non-adherent [14]. We will use pre- and post-randomization prognostic factors that predict censoring and the outcome to derive stabilized inverse probability of censoring weights (IPCW) [14,15]. We will then use these weights to fit a Cox proportional hazards (PH) model or pooled logistic regression; the choice of model will depend upon whether we treat time as continuous or discrete.

Estimates based on IPCW can be quite sensitive to the model generating the weights. When selecting variables to include in the weight determining model we will use an approach outlined previously [15]. This approach includes examining the distribution of weights and modifying the model if there are extreme values.

2.5.1.4.2. As-treated analyses using propensity score methods. We will estimate the effect of treatment actually received, with propensity score adjustment used to minimize bias. For this analysis we will calculate the treatment actually received during the intervention period as average total daily intake of vitamin D from study capsules and off-study supplements. For a person who neither dies nor experiences the event of interest during the intervention period, the average will be calculated over 5 years. Otherwise it will be calculated over the time from randomization until the event (or death).

We will define a person as 'treated' if their average total intake is $\geq$ 1600 IU/day, and 'not treated' if their average intake is $<$ 1600 IU/day, regardless of the group to which they were randomized. This threshold corresponds to a participant in the active arm taking 80% of their study tablets, without additional off-study supplementation.

We will use a generalized boosted model (GBM) [16] to estimate the propensity score for being treated, from which we will derive inverse probability of treatment weights (IPTW) [17]. The weights will be used in a weighted Cox PH model that regresses survival against an indicator of whether or not the person was treated. We will use bootstrapping to construct a 95% CI for the estimated HR [18].

The GBM will include baseline covariates that are associated with the outcome or confound the relationship between treatment and outcome [19]. We will use subject-matter knowledge to construct directed acyclic graphs from which we will identify the potential confounders for which adjustment is necessary. We will follow the recommendations from a paper on 'best practice' when using IPTW [20] to assess whether we have achieved balance between distributions of baseline covariates in the treated and untreated groups in the weighted sample. We will use standardized differences to compare proportions (of binary variables), means and higher-order moments. For continuous variables, we will also produce side-by-side boxplots and empirical cumulative distribution functions. If these diagnostics suggest that the groups are not sufficiently balanced in the weighted sample, then we will modify the GBM by including additional covariates. The process of deriving IPTWs and assessing balance will be repeated until we are satisfied that balance has been achieved.

We will conduct a second as-treated analysis in which the treatment received is categorized as 'very low' (average intake $<$ 200 IU/day); 'low' (average intake 200–499 IU/day); 'moderate' (average intake 500–1600 IU/day); and 'high' (average intake $>$ 1600 IU/day). Propensity score methods generalize relatively easily to this situation where treatment has more than two levels [21], and this analysis will allow us to assess whether or not vitamin D is associated with an outcome when taken at doses lower than prescribed by our protocol.

2.5.1.4.3. Rank-preserving structural failure time model. In a third approach, we will estimate the causal effect of vitamin D supplementation by fitting a rank-preserving structural failure time model (RPSFTM). This randomization-based method allows for participants switching between treatments more than once [22]. By classifying a participant as being 'on' the treatment whenever their total intake of vitamin D is $\geq$ 1600 IU/day and otherwise 'off' treatment, we will be able to partition a participant's observed event time (T_i) into time 'on' and 'off' treatment (i.e. $T_i = T_i^{on} + T_i^{off}$). We will relate T_i to the counterfactual event time (U_i) that would have been observed had participant i never been treated, using

$$\boldsymbol{U_i} = \boldsymbol{T_i^{off}} + \boldsymbol{T_i^{on}} \exp(\boldsymbol{\psi}). \quad (1)$$

here ψ is the true causal parameter (i.e. the treatment effect) and $\exp(\psi)$ is the acceleration factor that describes the relative increase or decrease in survival due to treatment. We will use a grid search based upon the Wald test from a Cox PH model adjusted for randomization strata to estimate the value of ψ that balances U_i across randomization groups and we will estimate a 95% CI for ψ [23].

2.5.1.5. Hypothesis tests, significance levels, multiple testing. Hypothesis tests, when performed, will be two-sided and, unless otherwise specified, we will use a statistical significance level of $P < 0.05$. We will not adjust for multiple testing.

2.5.2. Predicting baseline 25(OH)D concentrations

We did not collect blood samples at baseline because the study was designed to replicate what would happen in a population subjected to fortification and we could not justify the additional cost (estimated at several millions of dollars). Hence, to be able to investigate whether or not the effect of supplementation is modified by a person's vitamin D status we need to predict baseline 25(OH)D concentrations. To do this we will use data and blood samples collected from the placebo group for compliance monitoring as training data to develop a prediction model as follows.

2.5.2.1. Preparing the training data. To model the seasonal component of 25(OH)D concentrations, we will fit the following sinusoidal model:

$$\textbf{actual } 25(\textbf{OH})\textbf{D concentration} = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \sin\left(\frac{2\pi \boldsymbol{t}}{12}\right) + \boldsymbol{\beta}_2 \cos\left(\frac{2\pi \boldsymbol{t}}{12}\right), \quad (2)$$

where t is the month of sample collection, and β_0, β_1 and β_2 are model coefficients to be estimated [24]. Deseasonalized 25(OH)D concentrations will then be estimated by adding the overall mean 25(OH)D concentration to the residuals from the fitted sinusoidal model.

2.5.2.2. Prediction modelling. We will use boosted regression trees (BRTs) [25–27] to model the relationship between continuous deseasonalized 25(OH)D concentration and a set of fifteen explanatory variables. The latter will include factors that are correlated with cutaneous production of vitamin D (ambient ultraviolet radiation at the person's place of residence, skin type, time outdoors, physical activity), factors that have been shown to be strong determinants of 25(OH)D concentration in the elderly (total intake of vitamin D from diet and supplements, body mass index, self-reported health status) [28], and personal and lifestyle factors (age, sex, history of chronic diseases (diabetes, high blood pressure, high cholesterol), smoking behaviour, alcohol consumption, living arrangements). Dietary vitamin D intake will be estimated using responses to a 16-item food frequency questionnaire; the items include oily fish, meat, margarine (which is routinely fortified in Australia), eggs, cheese, and milk fortified with vitamin D.

We will fit BRT models for 36 combinations of parameter settings, using learning rates of 0.01, 0.005 and 0.001, tree complexities of 2, 3, 4 and 5, and bag fractions of 0.5, 0.6 and 0.7. For each combination, we will use 10-fold cross-validation to determine the optimal number of trees [27]. To identify the 'best' set of parameters we will use the cross-validation deviance. The optimal fitted BRT model will then be used to predict deseasonalized baseline 25(OH)D concentration for all trial participants using data from their baseline questionnaire.

2.5.3. Blinding

Blinding will be broken after all results (from primary and secondary analyses) have been generated. Secondary analyses require estimation of total vitamin D intake (excluding dietary intake) for each person. In theory, this requires knowledge of whether or not a person was randomized to vitamin D or placebo. However, we can avoid breaking the blinding by generating two versions of the results: one set will be based on the assumption that randomization group A received vitamin D; the other set will assume that randomization group A received placebo. Once blinding has been broken, we will know which results are the 'true' estimates to be included in the final manuscript/s.

We will initially perform the analyses using a dataset for which the true randomization allocation has been replaced with a randomly assigned allocation. This will allow the statistician to test and validate all statistical programs. Once validated, a researcher external to the D-Health team will execute the programs to produce results for the actual randomization allocation; blinding will then be broken.

2.6. Tertiary outcomes: details of analyses to be reported in separate publications

Due to the long period of passive follow-up, some tertiary outcomes will be analyzed prior to the analysis of the primary and secondary outcomes. Detailed statistical analysis plans will be written prior to each analysis, but here we highlight some important issues.

To maintain blinding, any analysis performed prior to study completion will use a random sample of 8,000 participants from each treatment arm. A different sample will be generated for each outcome (or group of related outcomes). A researcher external to the study team will randomly assign each person in the sample a new unique identifier, and the randomization code will be randomly mapped from its current values (A and B) to 0 or 1. The secretary of the data safety monitoring board (DSMB) will hold the mapping link. After results have been finalized the DSMB secretary will provide the study group allocation to enable interpretation of the results. The choice of sample size is pragmatic; it is small enough to make it almost impossible to deduce from participant characteristics whether the group that received vitamin D was randomization group A or B, but large enough that there should be adequate statistical power for most tertiary outcomes.

Interactions and subgroup analyses may differ from those considered for the primary and secondary outcomes, with factors chosen on the basis of their relevance to each specific outcome. Some of these factors, and indeed outcomes, will have missing data. The amount and nature of missing data in each random sample will inform which technique we use to handle missing data. For example, when there are very few missing data, we may use complete cases only, but we will also consider using multiple imputation and maximum likelihood estimation where there is a large amount of missing data. If we suspect data are missing not at random we will perform sensitivity analyses using pattern mixture models [29].

3. Discussion

We are publishing a detailed statistical analysis plan for the primary and secondary outcomes of the D-Health Trial while the intervention is on-going. This will help minimize bias during the analysis phase.

We have decided to use a FPSM for our primary analysis rather than Cox PH regression because the FPSM can accommodate non-proportional hazards and model differences over time in a continuous fashion. This is important because it is plausible that vitamin D supplementation would take time to have an effect on the primary and secondary outcomes, violating the PH assumption. The method also allows us to produce smooth plots of survival functions and the HR over time, providing greater insight into the treatment effect than is possible with a single 'average' HR from a Cox PH model [13].

Unlike many clinical trials, the D-Health Trial is population-based, of long duration, uses an intervention that can be readily obtained by participants, and ascertains many tertiary outcomes prior to the primary and secondary outcomes. Furthermore, access to vitamin D testing means that some participants may become unblinded to their

allocation, possibly influencing adherence to protocol. Indeed, this appears to be the case for VITAL participants; after 5 years, 10.8% of people in the placebo group were taking > 800 IU/day of off-study vitamin D supplements, compared with 6.4% in the vitamin D group [7]. There are thus considerable challenges to: (1) estimating the effectiveness of vitamin D treatment; and (2) maintaining investigator and analyst blinding through to the study's completion.

We have addressed the first challenge by proposing sophisticated secondary analyses that consider non-compliance and contamination. These approaches are appealing for a number of reasons. Throughout the intervention we collect data related to adherence and many prognostic factors that may affect it. We are therefore able to perform a per-protocol analysis that will appropriately adjust for post-randomization confounding and selection bias [30,31]. The other two methods (as-treated analysis incorporating propensity score methods and RPSFTM) do not censor a participant when they deviate from the protocol. Rather, they accommodate participants whose supplementary intake changes repeatedly over the course of the study. This type of treatment switching is not permitted by approaches that assume all-or-none compliance [32–35] or that participants in the placebo arm have no access to the active treatment [34]. Another attractive property of these methods is that we preserve some of the benefits of randomization; namely, that aside from the treatment, the two groups are balanced across measured and unmeasured confounders. The RPSFTM is randomization-based, while the as-treated analysis will mimic a randomized trial by creating a synthetic sample in which treated and untreated participants have similar distributions of measured baseline covariates [36].

Our proposed secondary analyses have some limitations. Unlike the FPSM, these analyses assume proportional hazards and each produces a single estimate of the effect. The validity of the per-protocol and 'as-treated' analyses depends upon there being no unmeasured confounders. A limitation of the RPSFTM is that the effect of treatment is assumed to be the same regardless of when it is received [23]. Moreover, to fit an RPSFTM to the actual event times we must make an assumption about participants' supplement intake after they have left the trial.

Our approaches to the second challenge, maintaining blinding, are designed to safeguard against bias until all analyses of the primary and secondary outcomes are finalized. This includes restricting analyses of tertiary outcomes to a subset of participants. Although this exclusion of participants is contrary to the intention-to-treat principle, it should not materially affect estimates because participants will be sampled at random.

It is likely that the effects of vitamin D supplementation are not linear and will be most marked in people who are vitamin D deficient prior to supplementation. Thus there have been recommendations that low 25(OH)D concentration be an inclusion criterion of vitamin D supplementation trials [37–39]. The D-Health Trial was not established to investigate the benefits of correcting for vitamin D deficiency. Instead we aim to determine whether increasing the mean 25(OH)D concentration in the population would result in substantial health benefits. As such, the trial aims to provide evidence to inform policy decisions regarding population-level interventions, such as mandatory food fortification.

The D-Health Trial uses a vitamin D dose equivalent to that used in VITAL. VITAL found that 2000 IU/day was sufficient to increase mean 25(OH)D concentrations in the supplemented group from 74 nmol/l at baseline to 104 nmol/l after one year, with no increased risk of adverse events during follow-up [7]. We are thus confident of achieving a substantial shift in 25(OH)D concentrations. Unlike VITAL, the D-Health Trial uses a monthly bolus dosing regimen and we expect higher compliance with the study tablets; approximately 80% of VITAL participants took at least two thirds of the trial capsules. Concerns have been raised about the efficacy [40] and safety [41] of bolus doses. Comparisons between D-Health and VITAL will enable us to assess this issue.

Publication of this statistical analysis plan in advance of the intervention's completion, and adherence to it, will avoid data-driven analyses of the primary and secondary outcomes and ensure robust reporting of outcomes. We have carefully considered how to handle the complexities of analyzing the D-Health trial; the methods we have outlined may prove useful to other researchers planning analyses of similar trials.

Trial status

Randomization of participants was completed in May 2015. Participants are currently completing their fourth or fifth year of the intervention.

Abbreviations

25(OH)D: 25-hydroxy vitamin D; BMI: body mass index; BRT: boosted regression trees; CI: confidence interval; DSMB: data safety monitoring board; FPSM: flexible parametric survival model; GBM: generalized boosted model; HR: hazard ratio; IPCW: inverse probability of censoring weights; IPTW: inverse probability of treatment weights; IU: international units; MBS: Medicare Benefits Scheme; PBS: Pharmaceutical Benefits Scheme; PH: proportional hazards; RPSFTM: rank preserving structural failure time model; VITAL: Vitamin D and Omega-3 Trial.

Grants and fellowships

The D-Health Trial is funded by a National Health and Medical Research Council project grant (APP1046681). RE Neale, DSA McLeod, PM Webb and DC Whiteman are supported by fellowships from the NHMRC.

Disclosures

The authors have nothing to disclose.

Acknowledgements

We would like to acknowledge the D-Health Trial staff. We also extend our thanks to the D-Health Trial participants who have committed to this research.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.conctc.2019.100333.

References

[1] P. Autier, M. Boniol, C. Pizot, et al., Vitamin D status and ill health: a systematic review, Lancet Diabetes Endocrinol 2 (1) (2014) 76–89.
[2] M. Gaksch, R. Jorde, G. Grimnes, et al., Vitamin D and mortality: individual participant data meta-analysis of standardized 25-hydroxyvitamin D in 26916 individuals from a European consortium, PLoS One 12 (2) (2017) e0170791.
[3] J. Lappe, P. Watson, D. Travers-Gustafson, et al., Effect of vitamin D and calcium supplementation on cancer incidence in older women: a randomized clinical trial, J. Am. Med. Assoc. 317 (12) (2017) 1234–1243.
[4] P. Autier, P. Mullie, A. Macacu, et al., Effect of vitamin D supplementation on non-skeletal disorders: a systematic review of meta-analyses and randomised trials, Lancet Diabetes Endocrinol 5 (12) (2017) 986–1004.
[5] J. Wactawski-Wende, J.M. Kotchen, G.L. Anderson, et al., Calcium plus vitamin D supplementation and the risk of colorectal cancer, N. Engl. J. Med. 354 (7) (2006) 684–696.
[6] R. Scragg, K.T. Khaw, L. Toop, et al., Monthly high-dose vitamin D supplementation and cancer risk: a post hoc analysis of the vitamin D assessment randomized clinical trial, JAMA Oncol (2018) e182178.
[7] J.E. Manson, N.R. Cook, I.M. Lee, et al., Vitamin D supplements and prevention of cancer and cardiovascular disease, N. Engl. J. Med. 380 (1) (2019) 33–44.

[8] G. Bjelakovic, L.L. Gluud, D. Nikolova, et al., Vitamin D supplementation for prevention of mortality in adults, Cochrane Database Syst. Rev. 1 (2014) Cd007470.
[9] R.E. Neale, B.K. Armstrong, C. Baxter, et al., The D-Health Trial: a randomized trial of vitamin D for prevention of mortality and cancer, Contemp. Clin. Trials 48 (2016) 83–90.
[10] Institute of medicine committee to review dietary reference intakes for vitamin D, calcium. The national academies collection: reports funded by national institutes of health, in: A.C. Ross, C.L. Taylor, A.L. Yaktine, H.B. Del Valle (Eds.), Dietary Reference Intakes for Calcium and Vitamin D, National Academies Press (US) National Academy of Sciences., Washington (DC), 2011.
[11] D. Moher, S. Hopewell, K.F. Schulz, et al., CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials, Int. J. Surg. 10 (1) (2012) 28–55.
[12] P. Royston, M.K. Parmar, Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects, Stat. Med. 21 (15) (2002) 2175–2197.
[13] P. Royston, M.K. Parmar, An approach to trial design and analysis in the era of non-proportional hazards of the treatment effect, Trials 15 (2014) 314.
[14] J.M. Robins, D.M. Finkelstein, Correcting for noncompliance and dependent censoring in an AIDS Clinical Trial with inverse probability of censoring weighted (IPCW) log-rank tests, Biometrics 56 (3) (2000) 779–788.
[15] S. Dodd, P. Williamson, I.R. White, Adjustment for treatment changes in epilepsy trials: a comparison of causal methods for time-to-event outcomes, Stat. Methods Med. Res. (2017) 962280217735560.
[16] D.F. McCaffrey, G. Ridgeway, A.R. Morral, Propensity score estimation with boosted regression for evaluating causal effects in observational studies, Psychol. Methods 9 (4) (2004) 403–425.
[17] P.R. Rosenbaum, Model-based direct adjustment, J. Am. Stat. Assoc. 82 (398) (1987) 387–394.
[18] P.C. Austin, Variance estimation when using inverse probability of treatment weighting (IPTW) with survival analysis, Stat. Med. 35 (30) (2016) 5642–5655.
[19] P.C. Austin, P. Grootendorst, G.M. Anderson, A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a Monte Carlo study, Stat. Med. 26 (4) (2007) 734–753.
[20] P.C. Austin, E.A. Stuart, Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies, Stat. Med. 34 (28) (2015) 3661–3679.
[21] D.F. McCaffrey, B.A. Griffin, D. Almirall, et al., A tutorial on propensity score estimation for multiple treatments using generalized boosted models, Stat. Med. 32 (19) (2013) 3388–3414.
[22] J.M. Robins, A.A. Tsiatis, Correcting for non-compliance in randomized trials using rank preserving structural failure time models, Commun. Stat. Theor. Methods 20 (8) (1991) 2609–2631.
[23] A. Allison, I.R. White, S. Bond, Rpsftm: an R package for rank preserving structural failure time models, R J 9 (2) (2017) 342–353.
[24] I.A. van der Mei, A.L. Ponsonby, T. Dwyer, et al., Vitamin D levels in people with multiple sclerosis and community controls in Tasmania, Australia, J. Neurol. 254 (5) (2007) 581–590.
[25] J.H. Friedman, Greedy function approximation: a gradient boosting machine, Ann. Stat. 29 (5) (2001) 1189–1232.
[26] T. Hastie, J.H. Friedman, R. Tibshirani, The Elements of Statistical Learning : Data Mining, Inference, and Prediction, Springer, New York, 2001.
[27] J. Elith, J.R. Leathwick, T. Hastie, A working guide to boosted regression trees, J. Anim. Ecol. 77 (4) (2008) 802–813.
[28] B. Tran, B.K. Armstrong, K. McGeechan, et al., Predicting vitamin D deficiency in older Australian adults, Clin. Endocrinol. 79 (5) (2013) 631–640.
[29] B. Ratitch, M. O'Kelly, R. Tosiello, Missing data in clinical trials: from clinical assumptions to statistical analysis using pattern mixture models, Pharmaceut. Stat. 12 (6) (2013) 337–347.
[30] M.A. Hernan, S. Hernandez-Diaz, J.M. Robins, Randomized trials analyzed as observational studies, Ann. Intern. Med. 159 (8) (2013) 560–562.
[31] M.A. Hernan, J.M. Robins, Per-protocol analyses of pragmatic trials, N. Engl. J. Med. 377 (14) (2017) 1391–1398.
[32] S.G. Baker, Analysis of survival data from a randomized trial with all-or-none compliance: estimating the cost-effectiveness of a cancer screening program, J. Am. Stat. Assoc. 93 (443) (1998) 929–934.
[33] J. Cuzick, P. Sasieni, J. Myles, et al., Estimating the effect of treatment in a proportional hazards model in the presence of non-compliance and contamination, J R Stat Soc Ser B 69 (2007) 565–588.
[34] T. Loeys, E. Goetghebeur, A causal proportional hazards estimator for the effect of treatment actually received in a randomized trial with all-or-nothing compliance, Biometrics 59 (1) (2003) 100–105.
[35] H. Nie, J. Cheng, D.S. Small, Inference for the effect of treatment on survival probability in randomized trials with noncompliance and administrative censoring, Biometrics 67 (4) (2011) 1397–1405.
[36] P.C. Austin, The use of propensity score methods with survival or time-to-event outcomes: reporting measures of effect similar to those used in randomized experiments, Stat. Med. 33 (7) (2014) 1242–1258.
[37] R.P. Heaney, Guidelines for optimizing design and analysis of clinical studies of nutrient effects, Nutr. Rev. 72 (1) (2014) 48–54.
[38] W.B. Grant, B.J. Boucher, Randomized controlled trials of vitamin D and cancer incidence: a modeling study, PLoS One 12 (5) (2017) e0176448.
[39] W.B. Grant, B.J. Boucher, H.P. Bhattoa, et al., Why vitamin D clinical trials should be based on 25-hydroxyvitamin D concentrations, J. Steroid Biochem. Mol. Biol. 177 (2018) 266–269.
[40] R. Jorde, G. Grimnes, Serum cholecalciferol may be a better marker of vitamin D status than 25-hydroxyvitamin D, Med. Hypotheses 111 (2018) 61–65.
[41] H.A. Bischoff-Ferrari, B. Dawson-Hughes, E.J. Orav, et al., Monthly high-dose vitamin D treatment for the prevention of functional decline: a randomized clinical trial, JAMA Intern Med 176 (2) (2016) 175–183.