Statistical Analysis Plan **December 21st, 2023**

NCT01817166

PHRC-N/2012/ET-01

Efficacy of Cholecalciferol (Vitamin D3) for Delaying the Diagnosis of MS After a Clinically Isolated Syndrome

Final detailed statistical analysis plan drafted before data lock and before the commencement of statistical analyses.

An initial data analysis will provide a description of each arm. Statistical results will be presented using mean and standard deviation for quantitative variables with a Gaussian distribution, and median and interquartile range for other quantitative variables. For qualitative variables, frequencies and associated percentages will be presented.

Demographic, clinical (initial relapse), radiological (diagnostic MRI), and biological (lumbar puncture, vitamin D) data will be described. In particular, it will be determined for each patient whether the diagnostic criteria of McDonald 2017 are met at inclusion (Thompson et al., 2018).

The statistical analysis will be conducted by the BESPIM of the CHU de Nîmes using R 4.1.1 or a later version (R Core Team (2021). R Foundation for Statistical Computing, Vienna, Austria).

Primary Analysis (Rate of EDA)

The primary objective is to assess the efficacy and tolerance over 2 years of treatment with Cholecalciferol (Vitamin D3) in patients who have experienced a clinically isolated syndrome at high risk of multiple sclerosis (CIS, according to the McDonald criteria of 2005). This efficacy will be measured by comparing the risk of conversion to relapsing-remitting multiple sclerosis (RRMS) within 2 years after CIS, according to 2005 McDonald criteria (new lesions on follow-up brain and spinal cord MRIs or new relapses) (McDonald et al., 2001; Polman et al., 2005), between the arm receiving Cholecalciferol and the arm receiving the placebo.

Following the actual diagnostic criteria for MS (McDonald 2017), most patients present with a diagnosis of active MS and conversion corresponds to the current concept of EDA (evidence of disease activity defined by clinical relapses or imaging (gadolinium-enhancing lesions or new or unequivocally enlarging T2/FLAIR lesions) (Lublin et al., 2013) an “no conversion” corresponds to NEDA (no evidence of disease activity). Hence, the event will be the occurrence of EDA within the two years of follow-up (relapse and/or appearance of new or enlarging lesions (NEL) on follow-up MRIs at 3, 12, and 24 months).

Patients lost to follow-up before experiencing EDA will be considered as not having had the event. A sensitivity analysis with the maximum bias assumption will be conducted by considering missing data (discontinuation of follow-up before the end of the 2-year follow-up with NEDA) in the interventional arm as failures (non-responders = EDA) and those in the control arm as successes (responders = NEDA).

The rate of EDA in both arms will be estimated and then compared between the two treatment arms using mixed-effects logistic regression (random center effect). The treatment arm will be the factor of interest in the model. Adjustment will be made for randomization stratification factors (center, contrast enhancement on diagnostic brain and spinal cord MRI), known prognostic factors listed in section 4.1.4 of the protocol (Age, Gender, Presence of contrast enhancement on the reference MRI, Lesion load in FLAIR in mm^3, Number of lesions with contrast enhancement on the reference MRI, EDSS score at baseline), and the time between CIS and the first administration of the treatment, which varies among patients (from a few days to 95 days = maximum of 90 days between CIS and the reference MRI + a maximum of 5 days between the reference MRI and the first administration of the study treatment, which takes place on the first Saturday post-MRI). Adjustment for the center effect will be performed by considering the center as a random effect in the model (Kahan 2014, Kahan et al. 2015, Edgar et al. 2021).

The number of lesions with contrast enhancement on the reference MRI could not be determined quantitatively. This information is provided by recording the presence of contrast enhancement with a precision level in 3 modalities (similar to the diagnostic MRI): no lesions with contrast enhancement, one lesion with contrast enhancement, and at least two lesions with contrast enhancement. This information will be used as an adjusting variable only if it is not significantly associated with contrast enhancement on the diagnostic MRI (a randomization stratification factor), for which an adjustment is performed. The statistical significance of the association will be assessed using the Chi-square test or Fisher's exact test if the Chi-square test is not applicable, with a significance level set at < 0.05.

Furthermore, the lesion load in FLAIR in mm^3 (reference MRI) is available only for 258 out of 316 patients (those for whom the MRI images could be retrieved). Moreover, obtaining this volume is not reliable because the software does not provide a sufficiently precise estimate. Therefore, no adjustment will be made for the lesion load, and the number of T2/FLAIR lesions on the reference MRI will be used instead as an indicator of lesion load (in 2 modalities: <9 lesions, >= 9 lesions).

In the event of the use of an incorrect stratum during randomization (randomization performed for the stratum with contrast enhancement on the diagnostic MRI while the true value was an absence of contrast enhancement, and vice versa), the analyses will be adjusted for the contrast enhancement value on the diagnostic MRI (Yelland et al., 2022). The stratum error rate will be described by arm.

<u>Primary Analysis (Time to EDA)</u>

Time to EDA is the duration between the date of the first administration of the experimental treatment and the date of EDA.

Date of EDA is:
- The date of the "conversion MRI" for patients with EDA determined by the appearance of NEL on MRI without having a relapse during the follow-up,
- The date of the first symptoms of a relapse for patients with EDA determined by a new relapse without the appearance of NEL on MRI during the follow-up,
- The earlier of the two dates between the date of the first symptoms and the date of the MRI showing NEL for patients who have both a relapse and NEL on MRI during the follow-up.

Patients lost to follow-up or not showing EDA during the follow-up will be right-censored at the date of the last news. The endpoint date will be established for each patient at the date of the end-of-study visit conducted 24 months (2 years) after inclusion.

Time to EDA in both arms will be graphically represented using the Kaplan-Meier method.

The proportional risk of EDA in the experimental arm compared to the placebo arm will be estimated along with its 95% confidence interval using a random-effects Cox model (random center effect) after checking the proportional hazards assumption. Adjustment will be made for the same variables as those used when analyzing the rate of EDA.

Secondary Analyses

Secondary Objective A: Evaluate clinical efficacy on the number of relapses during the study period. The determination of the number of relapses is based on data from the EDA visit and the summary table of relapses based on the evaluation at Month 24. The number of relapses will be compared between the two arms using a non-parametric Mann-Whitney Wilcoxon test.

Secondary Objective B1: Describe the parameters of magnetic resonance imaging (brain and spinal cord MRI) at the "conversion MRI":
- Appearance of NEL on brain MRI in T2/FLAIR (including size increase) and/or on spinal cord MRI in T2 ("conversion MRI" Yes/No)
- New brain or spinal cord gadolinium-enhancing T1 lesions (active lesion(s) Yes/No)
- Lesion load in mm^3 of each brain MRI from the FLAIR sequence
- Total number of gadolinium-enhancing T1 lesions (categorical measure: none, one, at least two)

Secondary Objective B2: Describe the magnetic resonance imaging (brain and spinal cord MRI) parameters during follow-up:
- Normalized brain volume obtained from the 3D T1 sequence at different visits
- Change in global brain volume (mm3, from the 3D T1 sequence) between the baseline examination and the 2-year examination on one hand, and between the 3-month examination and the 2-year examination on the other hand, to limit the effect of pseudo-atrophy after the first relapse

Secondary Objective C: Evaluate efficacy on disability progression, measured by the EDSS score and Kurtzke functional scores (Kurtzke 1983; Noseworthy et al. 1990; Goodkin et al. 1992) at baseline and during follow-up, then considering scores at M0 and M24 only.

Secondary Objective D: Evaluate efficacy on cognitive abilities (PASAT) (Cutter et al. 1999; Fischer et al. 1999; Parmenter et al. 2006) at baseline and during follow-up, then considering scores at M0 and M24 only.

Secondary Objective E: Evaluate efficacy on quality of life (EQ5D, SF36, TLS-QoL10, and TLS-COPING10 questionnaires), fatigue (FSMC questionnaire), and anxiety/depression (HADS questionnaire) at baseline and during follow-up, then considering scores at M0 and M24 only.

For Objectives C, D, and E:

- A first mixed-effects regression model will be used to compare the evolution of the outcome measures (EDSS score for Objective C, PASAT score for Objective D, and quality of life scores for Objective E). The model variables will include:
  - Scores at different visits (outcome variables)
  - Treatment arm (explanatory variable of interest)
  - Center as a random effect (multiple patients per center)
  - Patient as a random effect (multiple measurements per patient)
  - Visit (temporal chronological variable)
  - Interaction between the treatment arm and the presence of a relapse during follow-up
- A second mixed-effects regression model will be used to compare the M24 scores of the outcome measures while adjusting for baseline scores (M0):
  - M24 score (outcome variables)
  - Treatment arm (explanatory variable of interest)
  - Center as a random effect (multiple patients per center)
  - M0 score (adjustment variable)

- Interaction between the treatment arm and the presence of a relapse during follow-up

Secondary Objective F: Evaluate treatment tolerance on:
a. The occurrence of AEs and SAEs
b. Pharmacological measurement of serum concentrations of 25(OH)D2 and 25(OH)D3
c. Measurement of serum calcium, urinary calcium, and renal function

The number of AEs and SAEs per patient will be described by arm.

For points b and c, a mixed-effects regression model will be used to compare the evolution of measurements during follow-up. The model variables will include:
- Measurements at different visits (outcome variable)
- Treatment arm (explanatory variable of interest)
- Center as a random effect (multiple patients per center)
- Patient as a random effect (multiple measurements per patient)
- Visit (temporal chronological variable)

Secondary Objective G: Initial formulation: Correlate the variation in clinical and imaging parameters with the initial level and changes in the levels of 25(OH)D2 and 25(OH)D3 under treatment.

Reformulated Objective G1: Study the link between the severity of impairment (parameters measured by the baseline MRI, such as the number of lesions, EDSS score) and the vitamin D level measured during pre-screening (25(OH)D2 and 25(OH)D3).

A mixed-effects regression model will be estimated. The model variables will include:
- Severity of impairment at inclusion (outcome variable). Several variables will be successively modeled: binary variables for MRI parameters (using a logistic regression model) and a quantitative variable for the EDSS score.
- Levels of 25(OH)D2 and 25(OH)D3 at pre-inclusion (explanatory variables of interest)
- Treatment arm (adjustment variable)
- Center as a random effect (multiple patients per center)

Reformulated Objective G2: Study the link between the occurence of EDA and vitamin D levels (25(OH)D2 and 25(OH)D3) during follow-up.

A mixed-effects regression model will be estimated. The model variables will include:
- Occurrence of EDA (relapse or MRI) (outcome variable)
- Treatment arm (adjustment variable)
- Center as a random effect (multiple patients per center)

- Patient as a random effect (multiple measurements per patient)
- Levels of 25(OH)D2 and 25(OH)D3 measured during follow-up (explanatory variables of interest)
- Visit (temporal chronological variable)

Objectives H, I, and J do not require statistical analyses and involve the establishment of biobanks.

Secondary Objective K: Estimate the discordance rate between the presence of NEL on follow-up MRI scans determined by the neurologist during the study and the result of the centralized MRI analysis performed at the end of the study (EDA not confirmed retrospectively or, conversely, detection of EDA not identified during the study). Additionally, estimate the rate of patients retrospectively identified as erroneously included based on centralized MRI reevaluation (inclusion criterion MRI not verified during centralized reading of the baseline MRI).

Discordance between decisions made during the study by neurologists and the results of centralized reevaluation analyses will be determined for each patient throughout his follow-up. The discordance rate will be estimated and presented with a 95% confidence interval.

The rate of patients retrospectively identified as erroneously included will also be estimated and presented with a 95% confidence interval.

Anticipated Level of Significance:

A difference will be considered statistically significant when the significance level of the test is less than or equal to 0.05.

Method for Handling Missing, Unused, or Invalid Data:

For the analysis of the primary outcome, if a patient is lost to follow-up without experiencing EDA, they will be considered as NEDA at the end of the 2-year follow-up and censored at the date of the last known status.

Regarding the adjustment variables in the primary analysis:

- If the gadolinium-enhancement status at the diagnostic MRI scan is unknown for a given patient, the missing data will be imputed by the gadolinium-enhancement status at the

baseline MRI scan for that patient if known. If the latter is also missing, the gadolinium-enhancement status used as the randomization stratification variable will be employed.

- When the date of the CIS is missing, it will be extrapolated from the date of the diagnostic MRI scan plus the median time between the CIS and the diagnostic MRI scan.

- If the number of T2/FLAIR lesions (used as an indicator of lesion burden, in two categories " <9 lesions " and " >= 9 lesions ") at the baseline MRI is missing, it will be analyzed in a third category "Number of Lesions Missing".

- If the EDSS score is missing at baseline, it will be imputed by the first non-missing EDSS scores at the M3 and M12 visits, subtracting the median of the EDSS score difference between M0 and M3 (respectively between M0 and M12) obtained from patients with scores recorded at both visits. If the result is negative, the median will not be subtracted.

For other secondary outcome measures, no method for imputing missing data is planned.

Analysis population:

Patients who did not receive the study treatment will not be considered in the analyses, which will be conducted on a modified intention-to-treat basis.

All patients who received at least one dose of the study treatment will be included in the analyses.

Thus, the following individuals will not be considered in the analyses:
- Patients with "Study discontinuation at the investigator's decision due to disease activity identified on the baseline MRI" or occurrence of a relapse before treatment onset.
- Patients who withdrew their consent.
- Patients who did not undergo a baseline MRI (withdrawal of consent, inability to undergo MRI).
- Patients lost to follow-up immediately after inclusion (at the inclusion visit and/or baseline MRI), with no follow-up data.
- Patients mistakenly included without MS or with occurrence of a relapse before the baseline MRI.

These patients cannot be included in an intention-to-treat analysis because they have no follow-up and, therefore, no information on disease activity during follow-up.

Bibliographic References

Edgar et al 2021 : Including random centre effects in design, analysis and presentation of multi-centre trials. Trials. 2021, 22:357.

Kahan 2014 : Accounting for centre-effects in multicentre trials with a binary outcome – when, why, and how? BMC Medical Research Methodology, 2014, 14:20.

Kahan et al 2015 : Many multicentre trials had few events per centre, requiring analysis via random-effects models or GEEs. J Clin Epidemiol. 2015 December ; 68(12): 1504–1511.

Lublin F.D. et al. Neurology. Defining the clinical course of multiple sclerosis: the 2013 revisions. 2014 Jul 15; 83(3): 278–286.

McDonald, W.I., Compston, A., Edan, G., Goodkin, D., Hartung, H.P., Lublin, F.D., McFarland, H.F., Paty, D.W., Polman, C.H. & Reingold, S.C. (2001) Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis. Annals of neurology, 50, 121-127.

Polman, C.H., Reingold, S.C., Edan, G., Filippi, M., Hartung, H.P., Kappos, L., Lublin, F.D., Metz, L.M., McFarland, H.F. & O'Connor, P.W. (2005) Diagnostic criteria for multiple sclerosis: 2005 revisions to the « McDonald Criteria ». Annals of neurology, 58, 840-846.

Thompson AJ, Banwell BL, Barkhof F, Carroll WM, Coetzee T, Comi G, Correale J, Fazekas F, Filippi M, Freedman MS, Fujihara K, Galetta SL, Hartung HP, Kappos L, Lublin FD, Marrie RA, Miller AE, Miller DH, Montalban X, Mowry EM, Sorensen PS, Tintoré M, Traboulsee AL, Trojano M, Uitdehaag BMJ, Vukusic S, Waubant E, Weinshenker BG, Reingold SC, Cohen JA. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. Lancet Neurol. 2018 Feb;17(2):162-173. doi: 10.1016/S1474-4422(17)30470-2. Epub 2017 Dec 21. PMID: 29275977 Review.

Yelland et al. 2023 : Handling misclassified stratification variables in theanalysis of randomised trials with continuous outcomes. Statistics in Medicine, 42(19):3529-3546.