

Statistical analysis plan

Section 1: Administrative Information

Title and trial registration: Melatonin for prevention of delirium in patients receiving mechanical ventilation in the intensive care unit: a multi-arm multi-stage adaptive randomized controlled clinical trial (DEMEL)

SAP version: version 1 dated

Protocol version: version 6.0 dated 2020.12.08

SAP contributors: Etienne Audureau

Signatures:

Senior statistician responsible: Etienne Audureau

Chief investigator/clinical lead: Keyvan Razazi

Section 2: Introduction

Background and rationale

Delirium is common in intensive care units (ICUs), affecting up to 80% of patients, and is particularly severe, associated with increased early and long-term mortality, higher costs, and severe cognitive impairment in survivors. It particularly affects sedated and ventilated patients and complicates and prolongs weaning from mechanical ventilation. The pathophysiology of delirium is complex and may involve circadian rhythms and sleep, as well as melatonin secretion, which are significantly altered in critically ill patients.

Melatonin secretion is central to the sleep-wake cycle. Previous studies suggest that melatonin is well tolerated, but the optimal dose to achieve a sufficient nocturnal peak and minimal morning concentration has not been determined. High doses may paradoxically disrupt the sleep-wake cycle.

Objectives and outcomes

The primary objective is to evaluate the pharmacokinetics and efficacy of melatonin in the prevention of delirium in ventilated patients in the ICU.

The final primary outcome measure for "efficacy" (phase 2) is the proportion of patients who develop delirium between randomization and Day 14 (or until death or discharge from ICU if these occur before Day 14).

The primary interim outcome measure for "activity" to select the most promising experimental arm during interim analysis is the percentage of patients with an "optimal" pharmacokinetic profile of melatonin 48 hours after the initiation of the study treatment.

Section 3: Study Methods

Trial design

This is a comparative, superiority, prospective, multicenter, randomized, controlled, double-blind, placebo-controlled, multi-arm, multi-stage (MAMS), phase II/III trial.

The study treatment is a single daily dose at a fixed time (9 PM) of 15 mL of an oral syrup, with one of the following three concentrations of melatonin: 0 mg/mL (0 mg/day, placebo); 0.02 mg/mL (0.3 mg/day, low dose of melatonin); or 0.2 mg/mL (3 mg/day, high dose of melatonin). All three dosages will be identical in appearance and volume. The 3 mg dose has been used in many therapeutic trials but may result in supraphysiological morning concentrations; thus, many authors have suggested the need to test lower, "physiological" doses, around 0.3 mg. The treatment will be administered until Day 14 (or until death or discharge from intensive care if these occur before Day 14).

We will use an adaptive MAMS methodology with 3 arms and 2 phases. The first phase, called the "activity" phase, aims to compare each of the 2 intervention arms with the same placebo group. At the end of this phase, recruitment is stopped in the less promising intervention arm based on the interim assessment criterion (pharmacokinetics), following a "pick the winner" strategy. Enrollment in the placebo group and the more promising intervention arm will continue into the next phase, called the "efficacy" phase, until the required number of patients to analyze the final adjudication criterion (based on delirium prevention) is reached. This design allows for the selection of the best experimental arm and significantly reduces the number of patients to be enrolled, as well as the cost and duration of the research, compared to a strategy involving two separate trials.

Randomization

The randomisation list is drawn up by the Clinical Research Unit (URC of the Henri Mondor University Hospitals) before the start of the trial, in balanced blocks and stratified according to hospital centre.

Sample size

Based on the available data on the pharmacokinetics of melatonin and its effect on the incidence of delirium, the interim primary endpoint, i.e. the proportion of patients with an "optimal" pharmacokinetic profile, is estimated to be <15% in the placebo group.¹ The final primary endpoint, the proportion of patients with delirium within 14 days of randomization, is estimated to be 32%.² The corresponding rates of these intermediate/final primary endpoints reasonably expected in the melatonin arm with the best pharmacokinetic profile (to be selected at the interim analysis) are 80%/16%, and those in the less promising arm are 60%/25%.^{3,4}

Using the method proposed by Friede et al⁵ and implemented by Parsons et al^{6,7} for designing adaptive trials, taking into account an overall alpha risk of 2.5% in a one-sided configuration, a correlation of 0.5 between the interim and final endpoints, complete follow-up of all patients up to the final endpoint, and enrollment of 25 patients in each arm for the interim analysis (75 total), and 135 for the final analysis (270 total ; 345 total for the trial) provided >80% power to reject the null hypothesis and to detect efficacy in one experimental arm. A total of at least 345 patients is therefore planned to be included. The calculations were performed by running 1000 simulations using the R package 'asd'.⁷

Statistical interim analyses and stopping guidance

The interim outcome measure for selecting the most promising treatment at the end of the first "activity" phase will be the percentage of patients with an "optimal" pharmacokinetic profile of melatonin 24 hours after the initiation of the study treatment.

For the interim analysis, only the treatment arm with a test statistic greater than zero and the highest of the two treatment arms will continue recruitment into the efficacy phase alongside the control arm.

Timing of final analysis

All outcomes analysed collectively

Timing of outcome assessments

Patients are followed up to 3 months

Section 4: Statistical Principles

Level of statistical significance

A two-tailed P-value of <0.05 indicated statistical significance.

Analysis populations

Primary endpoint analyses are performed according to randomization group on intention-to-treat (ITT) basis. The main analyses do not include patients with no follow-up information after early consent withdrawal or those excluded earlier for violation of eligibility criteria. Additional supportive analyses are performed on the per protocol (PP) population who did not deviate from the protocol, i.e. have complete information on the primary endpoint and a $\geq 75\%$ adherence to the initially allocated treatment from randomisation to day 14 (or until death or ICU discharge, whichever comes first).

Section 5: Trial Population

Screening data

Screening data will be collected and reported to describe representativeness of trial sample

Eligibility

Inclusion criteria

- Inclusion criteria
- Age ≥ 18 years ;
- Invasive mechanical ventilation;
- Expected length of stay in intensive care unit (ICU) greater than 48 hours
- Written informed consent (patient, next of skin, or emergency situation).

Exclusion criteria

- Invasive mechanical ventilation of more than 48 hours prior to randomization;
- Pregnancy and lactation;
- Hearing loss, non-French speaking;
- Severe dementia (Mini Mental State <20) or psychosis
- Delirium onset before or at randomization
- Alcohol withdrawal syndrome with Cushman score >5 before or at randomization
- No enteral route available, vomiting
- Severe liver failure (prothrombin time $<30\%$)
- Treatment with melatonin or medications that interact with or modify its metabolism (such as fluvoxamine, 5- or 8-methoxypsoralen, estrogen, cimetidine, carbamazepine, rifampicin);
- Moribund patient;
- Patients under guardianship or curatorship;
- Patient deprived of liberty;
- Patient not affiliated to social security;

- Patient already involved in another interventional study on melatonin or in another study with delirium as the primary objective;
- Hypersensitivity to melatonin;
- Admission to intensive care for cardiac arrest, stroke, traumatic brain injury or neurosurgery.

Section 6: Analysis

Outcome definitions

The primary final outcome measure (end of second phase: "efficacy") will be the percentage of patients who develop delirium between randomization and day 14 (or until death or ICU discharge if this occurs before day 14). In accordance with previous studies and recommendations, delirium will be assessed twice daily using the CAM-ICU (Confusion Assessment Method for the Intensive Care Unit) score,⁸ in the absence of coma (Richmond Agitation–Sedation Scale ≥ 3)⁹ and at least 2 hours after stopping sedatives.¹⁰

The primary interim endpoint for selecting the most promising treatment (end of first phase: "activity") will be the percentage of patients with an "optimal" melatonin pharmacokinetic profile 24 hours after initiation of study treatment. This profile will be assessed blindly during the administration of the second dose of study drug and will be based on the measurement of residual serum melatonin concentration before administration (9:00 p.m.), at peak (9:30 p.m.), and in the morning (8:00 a.m.) after administration. Based on previous pharmacokinetic studies, a profile is considered optimal if the peak concentration is >1000 pg/mL and the morning concentration is <100 pg/mL.³ These thresholds are consistent with the sensitivity of the technique used¹¹ and inter-individual variability and are calculated by subtracting the pre-dose baseline melatonin concentration to account for endogenous secretion, as suggested by previous pharmacokinetic studies of melatonin.¹²

Analysis methods

Descriptive statistical analyses were performed on the entire study population, particularly the randomised groups, to describe their general and baseline characteristics, demographics, medical history, and number of premature withdrawals. Quantitative variables are presented as means ($\pm$ standard deviation) or medians (25th -75th percentiles) according to the normality of their distribution, as assessed by Shapiro-Wilk tests and graphical methods. Qualitative variables are presented as percentages.

The final primary endpoint (at the end of the efficacy phase) was the percentage of patients who experienced delirium between randomisation and day 14 (or death or discharge from intensive care unit if this occurred before day 14). The interim primary endpoint for selection of the most promising dose regimen (at the end of the efficacy phase) was the percentage of patients with an optimal melatonin pharmacokinetic profile 24 hours after starting study treatment; this interim primary endpoint was chosen for its clinical relevance and high expected correlation with the final primary endpoint.

The study was to be stopped at the interim analysis if no treatment outperformed the control (test statistic <0). The final decision to stop the trial was to be made by the Steering Committee. Once 35 patients had been enrolled in one of the groups, the study was to be temporarily stopped while the interim analysis was being completed, if it had not already been completed.

Primary endpoint analyses were performed according to randomisation group on an intention-to-treat (ITT) basis. The main analyses did not include patients with no follow-up information after early

withdrawal of consent or those who were excluded earlier due to violation of eligibility criteria. Additional supportive analyses were performed on the per-protocol (PP) population that did not deviate from the protocol, i.e. had complete information on the primary endpoint and >75% adherence to the initially assigned treatment from randomisation until day 14 (or until death or hospital discharge, whichever came first).

For the analysis of the primary endpoints, the phase 2/3 MAMS design is based on a methodology that generalizes the Dunnett one-step test to a multistage model with a combination of tests, allowing overall control of the family-wise error rate (FWER), defined as the probability of rejecting ≥ 1 true null hypotheses, at a pre-specified level (2.5% one-sided). The approach described by Friede et al.⁵ was used, making use of all data collected during the two phases and the different randomisation arms. For the final analysis, the degrees of significance (p-values) at each stage of the study were obtained from Dunnett's statistical tests comparing the experimental arms with the control group on the final evaluation criterion; this approach allows the null hypothesis that all effects of the treatments compared with the control are equal to zero to be tested against the alternative hypothesis that at least one treatment effect is greater than zero. In order to use all available data for the final analysis, the p-values of the two stages were combined using a test combination approach¹³ based on the weighted inverse normal method.¹⁴

To account for the multiplicity of group/phase comparisons, the overall family-wise type I error rate (FWER) was controlled using a so-called "closed testing procedure",¹⁵ which allows overall control of the FWER in the strong sense, defined as the probability of rejecting ≥ 1 true null hypotheses. For the analysis of the interim endpoint, only the treatment group whose test statistic was greater than zero and was the highest between the two treatment arms continued recruitment into the efficacy phase together with the control group.

Secondary outcomes of randomized groups were compared using Chi square or Fisher's exact tests, according to expected numbers in crossings, for categorical variables, and using t-test or non-parametric Mann-Whitney test (pairwise comparisons), and ANOVA or Kruskal Wallis tests (comparisons of >2 groups) for quantitative variables, as appropriate. Effect sizes were reported as the absolute differences in proportions with asymptotic 95% CIs for binary endpoints (e.g., xxx) and as the median difference between treatments and corresponding 95% CIs calculated from Hodges-Lehman estimate for continuous endpoints (e.g., ventilator-free days). Analyses of time-to-event endpoints took into account the competing risks of hospital discharge (for mortality evaluation) and death or discharge (for time to first delirium). Cumulative incidence curves were plotted for each treatment group and compared using Gray test. Tolerance analysis examined the intervention-related adverse events according to their period of appearance and the concerned randomized group to compare rates and time of occurrence. A two-sided alpha risk of 5% was used for analyses of secondary outcomes.

Information on the primary endpoint was complete for the ITT analysis, while analyses of secondary endpoints were performed on a complete cases basis, with no imputation of missing information. All analyses were performed according to a predefined statistical analysis plan, using Stata v16.1 (StataCorp, College Station, TX, USA) and R 4.0.3 (R Foundation for Statistical Computing, Vienna, Austria). A two-tailed P-value of <0.05 indicated statistical significance.

Planned subgroup analyses

None

Missing data

No assumption to handle missing data.

Statistical software

Stata v16.1 (StataCorp, College Station, TX, USA) and R 4.0.3 (R Foundation for Statistical Computing, Vienna, Austria).

References

1. Bourne RS, Mills GH. Melatonin: possible implications for the postoperative and critically ill patient. *Intensive Care Med*. 2006;32(3):371-379. doi:10.1007/s00134-005-0061-x
2. Salluh JI, Wang H, Schneider EB, et al. Outcome of delirium in critically ill patients: systematic review and meta-analysis. *bmj*. 2015;350:h2538.
3. Vural EMS, van Munster BC, de Rooij SE. Optimal dosages for melatonin supplementation therapy in older adults: a systematic review of current literature. *Drugs Aging*. 2014;31(6):441-451. doi:10.1007/s40266-014-0178-0
4. Hatta K, Kishi Y, Wada K, et al. Preventive effects of ramelteon on delirium: a randomized placebo-controlled trial. *JAMA Psychiatry*. 2014;71(4):397-403. doi:10.1001/jamapsychiatry.2013.3320
5. Friede T, Parsons N, Stallard N, et al. Designing a seamless phase II/III clinical trial using early outcomes for treatment selection: An application in multiple sclerosis. *Statist Med*. 2011;30(13):1528-1540. doi:10.1002/sim.4202
6. Parsons N, Friede T, Todd S, et al. An R package for implementing simulations for seamless phase II/III clinical trials using early outcomes for treatment selection. *Computational Statistics & Data Analysis*. 2012;56(5):1150-1160. doi:10.1016/j.csda.2010.10.027
7. Parsons N. asd: Simulations for adaptive seamless designs. Published online November 26, 2013. Accessed February 20, 2016. <https://cran.r-project.org/web/packages/asd/index.html>
8. Ely EW, Inouye SK, Bernard GR, et al. Delirium in mechanically ventilated patients: validity and reliability of the confusion assessment method for the intensive care unit (CAM-ICU). *JAMA*. 2001;286(21):2703-2710.
9. Sessler CN, Gosnell MS, Grap MJ, et al. The Richmond Agitation–Sedation Scale: Validity and Reliability in Adult Intensive Care Unit Patients. *American Journal of Respiratory and Critical Care Medicine*. 2002;166(10):1338-1344. doi:10.1164/rccm.2107138
10. Patel SB, Poston JT, Pohlman A, Hall JB, Kress JP. Rapidly reversible, sedation-related delirium versus persistent delirium in the intensive care unit. *Am J Respir Crit Care Med*. 2014;189(6):658-665. doi:10.1164/rccm.201310-1815OC
11. Carter MD, Wade Calcutt M, Malow BA, Rose KL, Hachey DL. Quantitation of melatonin and n-acetylserotonin in human plasma by nanoflow LC-MS/MS and electrospray LC-MS/MS: Melatonin and N-acetylserotonin quantitation. *Journal of Mass Spectrometry*. 2012;47(3):277-285. doi:10.1002/jms.2051

12. Bourne RS, Mills GH, Minelli C. Melatonin therapy to improve nocturnal sleep in critically ill patients: encouraging results from a small randomised controlled trial. *Critical Care*. 2008;12(2):1.
13. Bauer P, Kieser M. Combining different phases in the development of medical treatments within a single trial. *Statist Med*. 1999;18(14):1833-1848. doi:10.1002/(SICI)1097-0258(19990730)18:14<1833::AID-SIM221>3.0.CO;2-3
14. Lehmacher W, Wassmer G. Adaptive sample size calculations in group sequential trials. *Biometrics*. 1999;55(4):1286-1290.
15. Stallard N, Friede T. A group-sequential design for clinical trials with treatment selection. *Statist Med*. 2008;27(29):6209-6227. doi:10.1002/sim.3436