nature portfolio

Peer Review File

Dronabinol for nightmares in post-traumatic stress disorder: a randomized controlled trial

Corresponding Author: Professor Stefan Roepke

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
General and specific comments to the manuscript NMED-A149202 submitted to Nat Med entitled „ Dronabinol for Nightmares in Post-Traumatic Stress Disorder “
This is a well-conducted, multicenter, double-blind, placebo-controlled, randomized controlled trial of the effects of dronabinol in 171 adults with PTSD and nightmares. The primary efficacy endpoint was the frequency and intensity of nightmares after the last intervention (10 weeks) using the CAPS-IV B2 instrument. The results showed a significantly greater reduction in CAPS-IV B2 scores with dronabinol (difference between groups: -1.49; 95% CI: -2.28 to -0.71; Cohen's d: 0.65; $P < 0.001$). Consistent with these results, some secondary endpoints also showed significant improvements in the dronabinol group (PSQI-A, PCL-5, ITQ, PGIC, EQ-5D-3L), while others did not reach significance (CAPS-5 total score, MADRS, EQ-5D-3L, SOFAS). Typical THC-related adverse events were reported.

Some additions could help to make the manuscript more complete:
1. There is conflicting information regarding the background of the study. On the one hand, it was pointed out that “dronabinol has not been evaluated” (abstract). On the other hand, the introduction reports that nabilone has reduced nightmares in patients with PTSD and that the study presented here aims to fill the gap in “robust evidence from larger, well-controlled trials of cannabinoid receptor agonists for the treatment of PTSD-related nightmares.” Mechanistically, both compounds (dronabinol, nabilone) work by activating cannabinoid receptors and thus mimic the effects of THC. Accordingly, it should be made clearer that dronabinol is a synthetic THC molecule, whose mechanism of action is linked to the activation of cannabinoid (CB) 1 and 2 receptors.
2. In connection with point 1, older data, such as that summarized in Passie et al. (2012; doi: 10.1002/dta.1377), should also be mentioned in the introduction. The studies cited there showed that changes in REM sleep can be observed equally with the administration of THC and THC-containing cannabis, and that in preclinical studies, the blocking of CB1 receptors prevents the effect of anandamide (the endogenous ligand on cannabinoid-receptors) on sleep. This suggests that the activation of CB1 receptors is an important mechanism in relation to sleep and sleep-related symptoms. In line with this, a retrospective cohort study (Casarett et al. 2019; doi: 10.1089/jpm.2018.0658) showed that increasing THC concentration is particularly beneficial in alleviating symptoms associated with neuropathic pain, sleep disorders, and depression, but not in treating flashbacks in PTSD or anxiety. This information may also be useful for discussing the results, particularly when considering why dronabinol did not provide relief for all symptoms of PTSD.
3. Participants were excluded if they had ever suffered from CUD or substance/alcohol dependence that had been diagnosed within the last 3 months. Please also indicate how many participants had ever used cannabis and for what reasons, e.g., to treat PTSD symptoms, and how the number of these study participants is distributed across the two treatment arms.
4. Regarding the placebo condition, it was stated that “There is no discernible difference between BX-1 (dronabinol) and the corresponding placebo with regard to color, taste, and appearance, thus ensuring the trials' double-blind character.” However, it remains unclear what kind of oily solution the participants assigned to the placebo arm received. Please clarify this. The same is true for dronabinol. In what kind of oily solution was dronabinol dissolved and what kind of other substances are in this solution?
5. It also remains unclear whether the conditions for taking the medication were standardized, i.e., whether BX-1 or the placebo was taken with food or on an empty stomach. This is important because the effects of cannabis-based medication can depend on this. Please explain.
6. There was conflicting information regarding the primary endpoint. In line 42 and 160f, it was stated that the primary

endpoint is change in nightmare frequency and intensity from baseline to week 10, while in your methods paper (doi.org/10.1186/s12888-023-04818-5) and in the supplemental material, the frequency after the last intervention (10 weeks) was the primary outcome ("The aim is to show that the score at 10 weeks is lower in the intervention group (I) than the control group (C)"). The sample size calculation seems to be based on the latter endpoint. This approach makes sense assuming a broad variability of nightmare frequency at baseline as stated in the sample size estimation. Otherwise, a much larger number of study participants would have been necessary. Please clarify this point and adjust the wording accordingly.
7. Accordingly to point 6, please revise Figure 2 A. You should show the course of the absolute numbers for CAPS IV B2 (0-8). In addition, Figures 2 B and C could provide more information if you show the absolute numbers for the MADRS and CAPS 5 results, even though the secondary endpoints focus on changes from baseline. In Figure 2 D, please add the same explanation of how you dichotomized the data as in Table 2. Please add this also in the Tables S5 and S 13.
8. The sample size calculation was based on the assumption that no more than 10% of patients would drop out. However, the dropout rates were over 18%. In accordance with the statistical analysis plan contained in the study protocol, a per-protocol (PP) analysis was performed (Table S5). However, the PP analysis may not be the ideal approach for a sensitivity analysis in a superiority study. In addition to the "missing not at random" (MNAR) assumption and the J2R method, additional imputations using the "baseline observation carried forward" (BOCF) and "last observation carried forward" (LOCF) methods could support the analysis.
9. Some of the data from the T21 project was analyzed with regard to the connection between PTSD and depression (Lynskey et al. 2024; doi: 10.1192/bjo.2024.13). The main findings of this study were that depression is common among patients taking cannabis-based medications (CBM) for the treatment of PTSD and that CBM was particularly effective in improving symptoms in patients who suffered from depression. With regard to the frequency of nightmares in the pre-specified subgroup analysis, you showed that dronabinol is superior to placebo in both groups (depression, no depression). Given the findings of Lynskey's article, it is recommended to investigate the influence of depression on the primary and secondary endpoints when administering dronabinol. Please also show the extent of symptom burden separately for depression/no depression.
10. Since the links between nightmares and psychopathologies have not yet been sufficiently researched in many respects, it may be unclear in individual cases whether the nightmares are related to PTSD. For example, it is also conceivable that nightmares are triggered by medication or sleep-related breathing disorders. For this reason, it would be desirable to thoroughly examine the medications used to determine whether medications that can cause nightmares could be a disruptive factor in the response to dronabinol therapy. Medications that are strongly associated with increased nightmares include beta-receptor blockers (e.g., propanolol, metoprolol, etc.), various antidepressants (e.g., SSRIs, SNRIs, tricyclics, etc.), and sedatives (e.g., zolpidem, melatonin, antihistamines, etc.). A subgroup analysis of the primary and secondary outcomes is recommended.
11. Sleep-related breathing disorders (e.g., obstructive sleep apnea syndrome) should also be reported. If this data cannot be collected retroactively, please indicate this potential bias in the Limitations section.
12. You stated that seven SAEs occurred in the dronabinol group, which were "all assessed as unrelated." These included tachycardia, which led to a 2-day hospitalization. Tachycardia is a typical side effect of dronabinol and CBM. Why was tachycardia assessed as "unrelated" in this case?
13. In the dronabinol group, no increase in cannabis withdrawal symptoms was observed at week 12 based on the total Marijuana Withdrawal Checklist (MWC). This is surprising insofar as a rebound phenomenon in REM sleep is known to occur after cannabis withdrawal, accompanied by an increase in the frequency and intensity of nightmares. I therefore recommend paying particular attention to the "sleep disorders" item on the MWC checklist. Please analyse, report and discuss this point in more detail in the Discussion.

Reviewer #2

(Remarks to the Author)
This manuscript was a multi-centre double blind placebo controlled trial of dronabinol for the treatment of nightmares in post-traumatic stress disorder. The primary outcome of the trial was significant, in that once daily dronabinol before bed for 10 weeks resulted in a significant reduction in the frequency and intensity of nightmares (CAPS IV B2) relative to placebo. There was also significant reduction in PGIC, while clinician assessed CAPS 5 overall scores and MADRS scores did not significantly differ between dronabinol and placebo. Importantly there were no serious adverse events. Overall, this is a well controlled and executed study and the findings are relevant and replicate and extend previous work from open label studies and underpowered RCTs. The comparative groups were appropriately balanced with respect to a host of criteria and the authors included, and recognized the importance, of a blinding assessment in the subjects given the psychoactive nature of the drug. I think this will be an important study for the field and will have significant influence on the focus of treatment of cannabinoid-based drugs for the management of specific PTSD symptoms, particularly sleep disturbances and nightmares.

I only have two points which require addressing.

1) I cannot find the baseline values for the CAPS IV B2, CAPS V (overall PTSD score) and MADRS, the only data presented are difference scores. The baseline data need to be included here, at least in table, to ensure that the two populations didnt have significant baseline differences in symptom severity. If there were any differences, I couldnt find any discussion within the statistics section as to whether these baseline values were included as covariates to control for any influence that could have had on the outcomes.

2) Perhaps this is a reflection of the nature of the question in the different versions, but why did the authors have the primary outcome measure be from CAPS IV but then the overall assessment of PTSD severity from CAPS 5? I couldnt find any mention of this within the methods section but would help to provide this information to understand the choice the outcome measures.

Reviewer #3

(Remarks to the Author)
The authors find that dronabinol is more effective than placebo for the treatment of nightmares in individual with posttraumatic stress disorder (PTSD).

The study builds upon prior, smaller, sometimes uncontrolled studies of cannabinoids (e.g., nabilone) for the treatment of PTSD-related nightmares or other PTSD symptoms. The size and rigor of this study are noteworthy advances over these prior works.

The trial was pre-registered (NCT04448808) and the design and intended outcomes have remained true to the original submission.

The writing is clear and cogent, and the manuscript is suitably succinct.

The finding that dronabinol was significantly superior to placebo on the primary outcome -- nightmare frequency and intensity -- is interesting and important. It is also collaterally supported by some of the secondary sleep-related outcomes, e.g., PSQI-A, where dronabinol was also superior to placebo. Of additional interest is the finding that overall improvement on the observer-rated CAPS-5 was not superior on dronabinol compared to placebo; but on the self-rated PCL-5, overall improvement was better on dronabinol compared to placebo. The authors may wish to discuss further this apparent discrepancy between the CAPS-5 and the PCL-5. Could it be that individuals with PTSD value improvement in sleep such that they feel they are overall improved, even if they do not experience broad-based improvement in PTSD symptoms?

There are a few issues that should be clarified:

1. If the CAPS-5 was used in the study, and the total score was used as one of the secondary outcomes, why was the primary (and some of the other secondary) outcome relying on the CAPS-IV B2 item? Why not the CAPS-5 B2 item?

2. Among the eight SAEs -- all of which occurred on drug -- two cases are referred to as "overdose of study drug". This clearly must be related to treatment, as it involves study drug, yet this and all the SAEs are reported as having been "not related to treatment". How is this possible? Was this an overdose with intent to harm oneself, overdose with intent to obtain additional effects of the drug, overdose in error? The nature and explanation of these two SAEs bears further explanation.

3. There is one primary outcome and twenty secondary outcomes as listed with the clinical trial registry (NCT04448808). The authors do a good job of presenting all (or most) of these in the study. But it would be important for the authors to take into account this number of comparisons, and justify how they evaluate statistical significance in the context of 20 secondary outcomes. I don't necessarily need to see the authors apply a post-hoc correction, but I would like to see some rationale provided to the reader as to how to interpret and value this multiplicity of secondary outcomes.

Reviewer #4

(Remarks to the Author)
The manuscript reports the results of a multicenter, double-blind, randomized, placebo-controlled trial of dronabinol (once daily before bedtime for 10 weeks) in adults with PTSD and nightmares. Overall, the results are clearly reported and largely adhere to established CONSORT and clinical trial reporting guidelines. The study addresses an important clinical question, and the statistical methods are generally appropriate. Below are several suggestions to improve clarity, transparency, and methodological rigor.

1. Toward the end of the “Trial Design and Oversight” subsection of the Methods section, please specify who had full access to the trial data, who performed the statistical analyses, and whether the authors had independent access to the data apart from the sponsor.
2. The trial used block randomization with a fixed block size of four. Please provide justification for using a fixed block size rather than varying block sizes, particularly given potential concerns about predictability in allocation.
3. The intention-to-treat (ITT) population was defined as all randomized patients who received at least one dose of study medication and had post-baseline data. Because this definition excludes randomized patients who did not receive treatment or did not provide post-baseline data, it does not meet the strict definition of ITT. Please either refer to this population as a modified ITT (mITT) population or justify why this definition is consistent with estimand and analytic framework of the trial.
4. The primary endpoint (change in CAPS-IV B2 score from baseline to week 10) was derived from a DSM-IV based instrument, whereas eligibility criteria were based on DSM-5. Please justify the choice of the DSM-IV B2 item as the primary outcome measure instead of using a DSM-5 based nightmare assessment and clarify the clinical rationale for this decision.
5. Missing data were handled using reference-based multiple imputation (jump-to-reference). Please clarify: (1) The underlying missing data assumption (e.g., MAR within treatment arm, treatment-policy estimand framework); (2) The rationale for choosing a reference-based (jump-to-reference) approach rather than standard MMRM alone; and (3) Whether the multiple imputation analysis constituted the primary analysis or a sensitivity analysis. Please also provide a brief explanation of the clinical and statistical rationale for this approach.
6. For the key secondary endpoints, the Hochberg procedure was used for multiplicity control. Please clarify whether this was implemented within a hierarchical testing strategy contingent on significance of the primary endpoint. Also, explicitly describe the order of testing and the family of hypotheses controlled.

7. In the Results section, please clarify why the primary analysis population (n = 169) excludes two randomized patients (n = 171 total randomized). Although the CONSORT diagram provides information, a brief explanation in the Results text would improve clarity and transparency.
8. The primary outcome was reported using Cohen's d. Please clarify whether this effect size was calculated using unadjusted means and pooled standard deviations or derived from the model-based (adjusted) estimates. If model-adjusted was used, please specify the covariates included in the model and list them in the corresponding table footnotes.
9. Given the evidence of partial functional unblinding, it is unclear whether any sensitivity analyses were conducted adjusting for treatment-assignment guess. If such analyses were performed, please report them. If not, please discuss whether this may have influenced patient-reported outcomes.
10. In Figure 2, please add the sample size at each visit for each treatment group. This would allow readers to better assess attrition over time and interpret longitudinal trends.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
I would like to thank the authors for their precise and transparent response to the comments on the first draft of their manuscript. The authors have provided a detailed response to my point 6 regarding the primary endpoint. To ensure that all readers have the opportunity to understand the authors' reasoning and rationale for deviating from the predefined plan to use ANCOVA and instead using MMRM, I would like to encourage the authors to explain this in more detail in the Statistics section. Please also indicate how the sample size was calculated according to the SAP. Furthermore, I would like to encourage the authors to include the comparison between MMRM and ANCOVA analyses in the appendix.

Reviewer #2

(Remarks to the Author)
The authors have addressed the minor points I had raised and requested and the revisions in accordance with the other reviewers requests as well has improved the manuscript.

Reviewer #3

(Remarks to the Author)
The authors have done a remarkable job of responding to copious -- and important -- reviewer critiques and suggestions. The resultant revision is clear, concise, and of considerable interest.

I have no recommendations for changes, other than to suggest to the authors that randomization does not guarantee that confounders are equally distributed between groups, especially in a study with this modest (though large for PTSD) sample size. They may want to temper some of the statements that they make in several places. For example: "...randomization should have balanced potential variability related to food intake across treatment arms." Bottom line: randomization reduces the risk of imbalances between groups on measured (and unmeasured) confounders, but it is not a panacea.

Reviewer #4

(Remarks to the Author)
The revised manuscript addresses most of my previous concerns; however, three issues remain unresolved and require further attention:

Previous Concern #2: "The trial used block randomization with a fixed block size of four. Please provide justification for using a fixed block size rather than varying block sizes, particularly given potential concerns about predictability in allocation."

The revised manuscript notes only that the block size was 4 but does not provide explicit justification for choosing a fixed rather than variable block size. Authors should directly address the predictability risk associated with a fixed block size of 4, as this is a recognized methodological concern in randomized controlled trials. Please include a clear rationale for this design choice.

Previous Concern #9: "Given the evidence of partial functional unblinding, it is unclear whether any sensitivity analyses were conducted adjusting for treatment-assignment guess. If such analyses were performed, please report them. If not, please discuss whether this may have influenced patient-reported outcomes."

The revised manuscript still does not include a sensitivity analysis for the primary outcome adjusted for treatment-assignment guess. If such an analysis was not conducted, this should be explicitly acknowledged. Authors should also provide justification for why it was not performed and discuss the potential implications of this omission for the interpretation of patient-reported outcomes.

Previous Concern #10: "In Figure 2, please add the sample size at each visit for each treatment group. This would allow readers to better assess attrition over time and interpret longitudinal trends."

This concern has not been addressed. Per-visit sample sizes for each treatment group must be included in Figure 2 to enable readers to appropriately assess attrition patterns and interpret longitudinal trends.

Dear Dr. Yang,

Thank you for inviting us to revise and resubmit our manuscript, entitled “Dronabinol for Nightmares in Post-Traumatic Stress Disorder". We appreciate the constructive criticism and suggestions we received from you and the reviewers.

We have now made major revisions to the manuscript and added numerous additional analyses including:

- Requested additional sensitivity analyses on potentially confounding factors such as co-medications
- Additional alternative imputation methods for the primary endpoint (LOCF, BOCF)
- Additional post hoc analyses on the CAPS-5 B2 item confirming the results on the pre-specified primary endpoint CAPS-IV B2
- Analysis of the primary endpoint based on an ANCOVA model comparing week-10 CAPS-IV B2 scores between treatment groups while adjusting for baseline CAPS-IV B2 score

Below, we have provided detailed point-by-point responses to the reviewers’ specific comments. All resulting changes made to the manuscript are highlighted in the revised version for easy reference. We adhered to the Sex and Gender Equity in Research (SAGER) guidelines throughout the reporting of this study.

As requested, we have reformatted the manuscript in accordance with the Nature Medicine Guide to Authors and corrected minor typographical errors (e.g., a typographical error in the abstract (between-group difference: reported as −1.49 instead of −1.50) was corrected to align with the manuscript). In accordance with your instructions, these structural changes to re-organize the manuscript are not tracked in the revised manuscript.

We hope that these revisions satisfactorily address all reviewer comments and meet the high standards for publication in *Nature Medicine*.

We look forward to your response.

Kind regards

Stefan Roepke (for the authors)

Editor´s general comment: The reviewers find the study is well designed. However, they have raised several technical concerns regarding the trial design, definition of endpoints (particularly the primary outcome) and whether the nightmares per se are related to PTSD.

Response: We thank the editors for the opportunity to revise the manuscript and for communicating the reviewers’ constructive comments. Detailed responses to each comment from the reviewers are provided in the point-by-point rebuttal below.

Editor’s specific comment: The editorial team have concerns regarding the unclear direction of research regarding Dronabinol for PTSD. Please in your response and in the discussion section of the manuscript discuss whether there are any larger clinical trials ongoing in testing the efficacy/effectiveness of THC/Dronabinol for PTSD.

Response: To our knowledge, apart from the present trial investigating dronabinol for PTSD-related nightmares, no registered clinical trials have specifically targeted nightmares in PTSD using oral THC/dronabinol or other cannabinoid agonists. A ClinicalTrials.gov search (search date: March 18, 2026; search terms: condition/disease “PTSD” and intervention/treatment “dronabinol” or “THC” or

"nabilone") identified 15 studies in total across all time and revealed a small number of ongoing or planned randomized studies investigating cannabinoid interventions in PTSD populations for which results have not yet been published. For example, NCT06222268 is a randomized trial examining THC, CBD, THC/CBD combinations and placebo administered prior to trauma-specific cognitive behavioural therapy (prolonged exposure) sessions in veterans with PTSD. Another registered study (NCT06381180) compares THC, CBD, THC/CBD and placebo over 12 weeks in individuals with PTSD. In addition, a larger randomized study of inhaled cannabis formulations in PTSD (NCT07224698; planned enrollment approximately 320 participants) is currently listed as not yet recruiting. Overall, while several cannabinoid-based PTSD trials are ongoing or planned, results from larger controlled studies are not yet available. In response to the editor's comment, we have now included an updated section on these ongoing trials in the Discussion section (see page 8, line 322): *"Several randomized clinical trials investigating cannabinoid-based interventions for PTSD are ongoing (e.g., NCT06222268, NCT06381180, NCT07224698), including studies using inhaled formulations and one evaluating cannabinoids as an adjunct to psychotherapy. However, none specifically target defined symptom domains, suggesting that a symptom-focused approach may represent a promising direction for future research."*

Reviewers' Comments:

Reviewer #1 (Remarks to the Author):

General comments: General and specific comments to the manuscript NMED-A149202 submitted to Nat Med entitled „ Dronabinol for Nightmares in Post-Traumatic Stress Disorder". This is a well-conducted, multicenter, double-blind, placebo-controlled, randomized controlled trial of the effects of dronabinol in 171 adults with PTSD and nightmares. The primary efficacy endpoint was the frequency and intensity of nightmares after the last intervention (10 weeks) using the CAPS-IV B2 instrument. The results showed a significantly greater reduction in CAPS-IV B2 scores with dronabinol (difference between groups: -1.49; 95% CI: -2.28 to -0.71; Cohen's d: 0.65; $P < 0.001$). Consistent with these results, some secondary endpoints also showed significant improvements in the dronabinol group (PSQI-A, PCL-5, ITQ, PGIC, EQ-5D-3L), while others did not reach significance (CAPS-5 total score, MADRS, EQ-5D-3L, SOFAS). Typical THC-related adverse events were reported. Some additions could help to make the manuscript more complete:

Response: Thank you very much, we have addressed the specific comments as detailed below. A typographical error in the abstract (between-group difference reported as −1.49 instead of −1.50) was corrected to align with the manuscript.

Specific comment 1. There is conflicting information regarding the background of the study. On the one hand, it was pointed out that "dronabinol has not been evaluated" (abstract). On the other hand, the introduction reports that nabilone has reduced nightmares in patients with PTSD and that the study presented here aims to fill the gap in "robust evidence from larger, well-controlled trials of cannabinoid receptor agonists for the treatment of PTSD-related nightmares." Mechanistically, both compounds (dronabinol, nabilone) work by activating cannabinoid receptors and thus mimic the effects of THC. Accordingly, it should be made clearer that dronabinol is a synthetic THC molecule, whose mechanism of action is linked to the activation of cannabinoid (CB) 1 and 2 receptors.

Response: Dronabinol is the International Nonproprietary Name (INN) for Δ9-tetrahydrocannabinol (THC) when used in a pharmaceutical context. In our study, dronabinol was not synthesized but

obtained from cannabis plant material through extraction and purification. Thus, dronabinol represents THC itself rather than a compound that mimics its effects. We removed the phrase "*efficacy of dronabinol … has not been evaluated*" from the abstract. Furthermore, we have clarified this point in the methods section to avoid potential confusion and to more clearly distinguish dronabinol (THC) from other cannabinoid receptor agonists such as nabilone, which is a synthetic cannabinoid analogue, see page 16, line 587: *"Dronabinol, the pharmaceutical form of Δ9-tetrahydrocannabinol (THC), was obtained from cannabis plant material by extraction and purification and dissolved in a stabilized triglyceride-based oil containing an antioxidant; the placebo consisted of the identical oil vehicle without dronabinol but included a cannabis flavoring to ensure indistinguishable appearance, taste, and smell."*

Specific comment 2. In connection with point 1, older data, such as that summarized in Passie et al. (2012; doi: 10.1002/dta.1377), should also be mentioned in the introduction. The studies cited there showed that changes in REM sleep can be observed equally with the administration of THC and THC-containing cannabis, and that in preclinical studies, the blocking of CB1 receptors prevents the effect of anandamide (the endogenous ligand on cannabinoid-receptors) on sleep. This suggests that the activation of CB1 receptors is an important mechanism in relation to sleep and sleep-related symptoms. In line with this, a retrospective cohort study (Casarett et al. 2019; doi: 10.1089/jpm.2018.0658) showed that increasing THC concentration is particularly beneficial in alleviating symptoms associated with neuropathic pain, sleep disorders, and depression, but not in treating flashbacks in PTSD or anxiety. This information may also be useful for discussing the results, particularly when considering why dronabinol did not provide relief for all symptoms of PTSD.

Response: We thank the reviewer for highlighting earlier mechanistic and observational evidence and elaborated these aspects in the revised version: We added Passie et al. (2012) to the Introduction to summarize data indicating that THC and THC-containing cannabis can alter sleep architecture/REM sleep and that CB1 receptor activation is implicated in sleep-related effects based on preclinical CB1 blockade findings. We also cite Casarett et al. (2019) in the Discussion to contextualize our domain-specific pattern of effects, noting that observational real-world data suggest THC concentration may be more strongly associated with improvements in sleep-related symptoms and depressed mood than with other PTSD-related symptoms, while emphasizing the non-randomized nature and different patient population of that cohort. Specifically, we now added to the manuscript: Introduction (Page 2, line 76): *"Consistent with this hypothesis, animal studies using exogenous cannabinoids, as well as experiments showing that anandamide-induced changes in sleep are abolished under CB1 receptor blockade, have demonstrated that CB1 receptor activation reduces wakefulness and arousal, promotes sleep, and increases the proportion of sleep spent in non–REM states (Passie et al., 2012…".* We also added a comment to the discussion (Page 6, line 245): *"In line with a primarily nocturnal dosing strategy and a sleep-focused mechanistic rationale, prior observational evidence also suggests that higher THC concentrations are more strongly associated with improvements in sleep disturbance and depressed mood than other PTSD symptoms (Casarett et al., 2019)."*

Specific comment 3. Participants were excluded if they had ever suffered from CUD or substance/alcohol dependence that had been diagnosed within the last 3 months. Please also indicate how many participants had ever used cannabis and for what reasons, e.g., to treat PTSD symptoms, and how the number of these study participants is distributed across the two treatment arms.

Response: We have clarified these issues in the revised manuscript. Lifetime Cannabis Use Disorder was an exclusion criterion in our trial. However, prior cannabis use per se and, if applicable, the reasons for cannabis use (e.g., self-medication of PTSD symptoms) were not systematically assessed

at baseline. We are therefore unable to provide specific data on previous cannabis exposure or its distribution across treatment groups. Having said that, randomization, however, prevents systematic differences across the treatment arms and protects against a potential confounding effect of previous cannabis exposure on the trial's results. In response to the reviewer's comments, we have now added a clarification to the Limitations section (*page 8, line 296*) to address this point. "*Prior cannabis use, including use for self-medication of PTSD symptoms, was not systematically assessed. While randomization reduces potential confounding by these factors, the trial does not allow conclusions regarding the extent to which prior exposure may have influenced treatment expectancy."*

Specific comment 4. Regarding the placebo condition, it was stated that "There is no discernible difference between BX-1 (dronabinol) and the corresponding placebo with regard to color, taste, and appearance, thus ensuring the trials' double-blind character." However, it remains unclear what kind of oily solution the participants assigned to the placebo arm received. Please clarify this. The same is true for dronabinol. In what kind of oily solution was dronabinol dissolved and what kind of other substances are in this solution?

Response: Both the active study medication (BX-1, dronabinol) and placebo were prepared using the same stabilized lipophilic oil vehicle containing an antioxidant. Dronabinol was dissolved in this oil in the active formulation, whereas the placebo contained the identical oil vehicle without dronabinol. To ensure indistinguishable sensory characteristics and maintain blinding, the placebo formulation additionally contained a cannabis flavouring. We have clarified the composition of the study medication and placebo in the Methods section (page 16, line 587). *"Dronabinol, the pharmaceutical form of Δ9-tetrahydrocannabinol (THC), was obtained from cannabis plant material by extraction and purification and dissolved in a stabilized triglyceride-based oil containing an antioxidant; the placebo consisted of the identical oil vehicle without dronabinol but included a cannabis flavoring to ensure indistinguishable appearance, taste, and smell."*

Specific comment 5. It also remains unclear whether the conditions for taking the medication were standardized, i.e., whether BX-1 or the placebo was taken with food or on an empty stomach. This is important because the effects of cannabis-based medication can depend on this. Please explain.

Response: Patients were instructed to administer the study medication (BX-1/dronabinol or placebo) as oil drops placed on a spoon and taken approximately one hour before bedtime. They were further instructed not to dilute the drops in water and not to take them together with liquids. Food intake at the time of administration was not standardized, and participants were not instructed to take the medication either with or without food. Because both treatment groups followed identical administration instructions and medication was taken at a consistent time in the evening, randomization should have balanced potential variability related to food intake across treatment arms.

We have clarified the administration procedure as follows: Methods (page 16, line 591): *"Patients were instructed to administer the study medication as oil drops on a spoon approximately one hour before bedtime and not to dilute the drops in water or to take them with liquids. Intake with or without food was not standardized."* Limitations/Discussion (page 8, line 309): *"Food intake at the time of medication administration was not standardized. Although food intake can influence the absorption of orally administered cannabinoids, identical administration instructions and randomization make systematic bias between treatment groups unlikely, with individualized dose titration further mitigating variability in exposure."*

Specific comment 6. There was conflicting information regarding the primary endpoint. In line 42 and 160f, it was stated that the primary endpoint is change in nightmare frequency and intensity

from baseline to week 10, while in your methods paper (doi.org/10.1186/s12888-023-04818-5) and in the supplemental material, the frequency after the last intervention (10 weeks) was the primary outcome ("The aim is to show that the score at 10 weeks is lower in the intervention group (I) than the control group (C)"). The sample size calculation seems to be based on the latter endpoint. This approach makes sense assuming a broad variability of nightmare frequency at baseline as stated in the sample size estimation. Otherwise, a much larger number of study participants would have been necessary. Please clarify this point and adjust the wording accordingly.

Response: We agree that the wording of the primary endpoint in the methods paper, the study protocol, and the current manuscript differed. The analysis model was revised from ANCOVA (methods paper and protocol) to MMRM in the SAP before database lock and unblinding. In the study protocol and methods paper, the primary analysis was described using an ANCOVA model comparing week-10 CAPS-IV B2 scores between treatment groups while adjusting for baseline CAPS-IV B2 score. In the SAP, the primary analysis was specified using a mixed model for repeated measures (MMRM) with change from baseline as the dependent variable and baseline CAPS-IV B2 score included as a covariate. This modification was chosen to allow the use of all available post-baseline observations and is consistent with recommended approaches for repeated-measures data in clinical trials (Detry et al., 2016).

Importantly, both modeling approaches estimate the treatment effect while adjusting for baseline values and therefore correspond to the same estimand. Accordingly, the sample size calculation based on a baseline-adjusted week-10 analysis is consistent with the MMRM-based primary analysis. Regulatory guidance indicates that when baseline is included as a covariate in a linear model, the estimated treatment effects are equivalent whether the analysis is based on the raw outcome or on change from baseline (EMA Guideline on adjustment for baseline covariates in clinical trials, Section 5.6). We additionally conducted an ANCOVA analysis comparing week-10 CAPS-IV B2 scores between treatment groups while adjusting for baseline CAPS-IV B2 score. The resulting treatment effect estimates were virtually identical (−1.495 for change from baseline using MMRM and −1.512 for the week-10 score using ANCOVA). In light of these findings, we did not include the additional analysis in the manuscript or Supplementary Appendix but would be happy to provide it upon request.

Specific comment 7. Accordingly to point 6, please revise Figure 2 A. You should show the course of the absolute numbers for CAPS IV B2 (0-8). In addition, Figures 2 B and C could provide more information if you show the absolute numbers for the MADRS and CAPS 5 results, even though the secondary endpoints focus on changes from baseline. In Figure 2 D, please add the same explanation of how you dichotomized the data as in Table 2. Please add this also in the Tables S5 and S 13.

Response: In response to the reviewer's comments, we have now restructured the depiction of the results: Baseline values for the primary and key secondary endpoints have now been added to Table 1. For the graphical presentation in the main manuscript, however, we retained the original version showing least squares (LS) means to present the prespecified mixed model for repeated measures (MMRM) analysis. This is based on the consideration that the primary analysis was based on change from baseline, so we feel that the figures of model-based estimates should display the change-from-baseline scale, with both treatment groups therefore start at zero, to be consistent with the statistical approach used.

To address the reviewer's suggestion and provide full transparency regarding the absolute outcome values, we now additionally provide the Supplementary Figures 2, 3, and 4 showing the observed raw values for the primary and key secondary outcomes at each visit in the Supplementary Appendix, together with the number of participants contributing data at each time point.

Regarding Figure 2D, we have added the same explanation of the dichotomization procedure as used in Table 2. The corresponding explanations have also been added to Tables S5 and S13.

Specific Comment 8. The sample size calculation was based on the assumption that no more than 10% of patients would drop out. However, the dropout rates were over 18%. In accordance with the statistical analysis plan contained in the study protocol, a per-protocol (PP) analysis was performed (Table S5). However, the PP analysis may not be the ideal approach for a sensitivity analysis in a superiority study. In addition to the "missing not at random" (MNAR) assumption and the J2R method, additional imputations using the "baseline observation carried forward" (BOCF) and "last observation carried forward" (LOCF) methods could support the analysis.

Response: To clarify, the original sample size calculation assumed that up to 10% of randomized patients might not receive study medication and therefore would not be included in the full analysis set, as specified in the Statistical Analysis Plan. To account for this conservative assumption, the required sample size of 158 patients was increased to 176 (88 per group). In the completed trial, 171 patients were randomized and 169 were included in the analysis; two patients (1.17%) were excluded according to the SAP (one did not receive study medication and one had no post-baseline assessment; see also figure 1, CONSORT flow diagram), which is well below the assumed 10%. The higher proportion referred to by the reviewer likely reflects treatment discontinuations after treatment initiation and therefore does not affect the assumptions underlying the sample size calculation.

The prespecified primary analysis was conducted in the full analysis set using the mixed model for repeated measures (MMRM). Missing data were handled using reference-based multiple imputation with a jump-to-reference (J2R) strategy as defined in the Statistical Analysis Plan. In this framework, patients who discontinue treatment are assumed to follow the outcome trajectory of the control group from the time of discontinuation onward, representing a conservative MNAR assumption in a superiority trial. The J2R approach typically yields conservative estimates in superiority trials because any treatment benefit is assumed to cease after treatment discontinuation.

In addition, the SAP prespecified sensitivity analyses, including a per-protocol analysis and an analysis based on the last CAPS-IV B2 assessment under treatment. Following the reviewer's suggestion, we performed additional imputations using BOCF and LOCF, as well as a complete-case analysis including only participants with observed outcome data at the relevant time point to assess robustness to different assumptions about missing data. The resulting treatment effect estimates were highly consistent with the primary analysis and were slightly larger in magnitude, further supporting the robustness of our findings. Alterations of the analysis set or imputation approach did not change the conclusions; in particular, treatment effects remained statistically significant across all analyses. These additional analyses are now included in Table S5 in the Supplement and are referenced in the manuscript as follows: "Sensitivity analyses yielded consistent results across all analysis sets (Table S5)" (see page 6, line 214).

Method	**estimate**	**95% CI**	**p value**
FAS	-1.50	-2.28 to -0.71	0.0002*
PP	-1.61	-2.60 to -0.63	0.0015
Last value under treatment carried forward	-1.58	-2.34 to -0.82	0.0001
Complete case	-1.67	-2.47 to -0.88	0.0001

LOCF	-1.55	-2.32 to -0.78	0.0001
BOCF	-1.53	-2.30 to -0.75	0.0001

* Only the p value for the FAS should be interpreted in a confirmatory manner

Specific comment 9. Some of the data from the T21 project was analyzed with regard to the connection between PTSD and depression (Lynskey et al. 2024; doi: 10.1192/bjo.2024.13). The main findings of this study were that depression is common among patients taking cannabis-based medications (CBM) for the treatment of PTSD and that CBM was particularly effective in improving symptoms in patients who suffered from depression. With regard to the frequency of nightmares in the pre-specified subgroup analysis, you showed that dronabinol is superior to placebo in both groups (depression, no depression). Given the findings of Lynskey's article, it is recommended to investigate the influence of depression on the primary and secondary endpoints when administering dronabinol. Please also show the extent of symptom burden separately for depression/no depression.

Response: We appreciate the reviewer's comment on the role of psychiatric comorbidities. The influence of depression on treatment outcomes was assessed in prespecified subgroup analyses (Figure 3), which did not indicate statistically significant heterogeneity of treatment effects.

To further address this point, we added descriptive summaries by baseline depression status. Participants with depression showed somewhat higher baseline symptom burden in CAPS-5 and MADRS (Supplementary Table 4). Descriptive change-from-baseline data (Supplementary Table 19) suggest that patients without depression experienced greater improvement, specifically in nightmare-related outcomes and PTSD symptoms than those with depression. However, depression did not significantly modify the treatment effect in the prespecified subgroup analysis. The following sentences were added: *"Descriptive baseline symptom burden by depression status is provided in Supplementary Table 4."* (page 3, line 120), "*Descriptive values for changes from baseline to week 10 in CAPS-IV B2, MADRS and CAPS-5 total scores by current depression status are presented in Supplementary Table 19.*" (page 6, line 221), and *"These findings do not confirm prior observational data suggesting greater benefit of cannabis-based medications in patients with PTSD and comorbid depression compared with those without depression (Lynskey et al., 2024)."* (page 7, line 257, discussion).

Table S4: Baseline primary and key secondary outcomes by current depression status[a]

Outcome	No depression N=101	Depression N=68
CAPS-IV B2[b]	6.78 (0.92)	6.69 (0.95)
CAPS-5[c]	36.6 (9.66)	44.97 (10.38)
MADRS[d]	16.1 (5.88)	23.29 (4.45)

[a] Data from the full analysis set. values are means (SD).
[b] The Clinician-Administered posttraumatic stress disorder (PTSD) Scale for DSM-IV (CAPS-IV) B2 item assesses frequency and intensity of nightmares; scores range from 0 to 8, with higher scores indicating greater symptom severity.
[c] CAPS-5 is the updated version for DSM-5 criteria assessing PTSD symptoms (range 0–80); with higher scores indicating greater symptom severity.
[d] The Montgomery–Åsberg Depression Rating Scale (MADRS) ranges from 0 to 60, with higher scores indicating more severe depression.

Table S19: Change from baseline to week 10 in primary and key secondary outcomes by current depression status[a]

Outcome	Dronabinol		Placebo	
	No depression	Depression	No depression	Depression
CAPS-IV B2[b]	−4.24 (2.52)	−2.87 (2.45)	−2.67 (2.43)	−1.56 (2.47)
N	45	31	42	27

Outcome	Dronabinol		Placebo	
	No depression	Depression	No depression	Depression
CAPS-5[c]	−11.55 (12.90)	−7.84 (13.72)	−6.17 (10.59)	−8.11 (10.75)
N	44	31	41	28
MADRS[d]	−3.91 (8.01)	−3.00 (8.13)	−1.40 (7.86)	−1.43 (8.21)
N	46	31	42	28

a Data from the full analysis set. values are means (SD). N=Patients with available data.
b The Clinician-Administered posttraumatic stress disorder (PTSD) Scale for DSM-IV (CAPS-IV) B2 item assesses frequency and intensity of nightmares; scores range from 0 to 8, with higher scores indicating greater symptom severity.
c CAPS-5 is the updated version for DSM-5 criteria assessing PTSD symptoms (range 0–80); with higher scores indicating greater symptom severity.
d The Montgomery–Åsberg Depression Rating Scale (MADRS) ranges from 0 to 60, with higher scores indicating more severe depression.

Specific comment 10. Since the links between nightmares and psychopathologies have not yet been sufficiently researched in many respects, it may be unclear in individual cases whether the nightmares are related to PTSD. For example, it is also conceivable that nightmares are triggered by medication or sleep-related breathing disorders. For this reason, it would be desirable to thoroughly examine the medications used to determine whether medications that can cause nightmares could be a disruptive factor in the response to dronabinol therapy. Medications that are strongly associated with increased nightmares include beta-receptor blockers (e.g., propanolol, metoprolol, etc.), various antidepressants (e.g., SSRIs, SNRIs, tricyclics, etc.), and sedatives (e.g., zolpidem, melatonin, antihistamines, etc.). A subgroup analysis of the primary and secondary outcomes is recommended.

Response: To examine medication with potential effect on dreaming and nightmares on treatment response, we have now conducted a comprehensive additional subgroup analysis comparing participants taking any such medication with those who were not and included this in the revised manuscript (Supplementary Table 17). Within the dronabinol group, the difference in treatment response between participants taking this medication and those not taking such medication was small and not statistically significant (estimate −0.20, p = 0.717). A similar pattern was observed in the placebo group (estimate −0.18, p = 0.752). Importantly, the treatment effect of dronabinol compared with placebo was similar in both strata. The estimated difference between dronabinol and placebo was -1.51 points in participants taking these medications and -1.48 points in those not taking such medication.

Taken together, these findings do not indicate that the use of medications potentially associated with nightmares meaningfully modified the treatment effect of dronabinol on the primary outcome. The new supplementary table also provides a detailed list of all medication with potential effect on dreaming and nightmares in this analysis, including their pharmacological classes and literature references supporting this classification (Supplementary Table 17, footnote). This additional analysis was conducted post hoc to explore a potential confounding factor and does not alter the conclusions of the trial's results. We added to the post hoc analysis subsection of the results part (page 6, line 217): *"Some permitted concomitant medications may affect dreaming and nightmares; however, analyses showed no impact on the primary outcome (Supplementary table 17)."*

Table S17. Impact of medication with potential effect on dreaming and nightmares* on treatment response

Contrast	Estimate	95% CI	p value
Effect of medication[a] (overall)	−0.02	−1.59 to 1.55	0.98
Effect of medication[a] within the dronabinol group	−0.20	−1.30 to 0.90	0.72
Effect of medication[a] within the placebo group	−0.18	−1.31 to 0.95	0.75
Treatment effect (dronabinol vs placebo) in patients taking medication[a]	−1.50	−2.73 to −0.28	0.016

Contrast	Estimate	95% CI	p value
Treatment effect (dronabinol vs placebo) in patients not taking medication[a]	−1.48	−2.49 to −0.47	0.004

*Medication was defined based on published reports linking specific pharmacological classes to increased dream intensity, vivid or abnormal dreams, and nightmares. The following drug classes and active substances used in the study were considered: lipophilic β-blockers (metoprolol); selective serotonin reuptake inhibitors (sertraline, escitalopram, fluoxetine, citalopram); serotonin–norepinephrine reuptake inhibitors (venlafaxine, duloxetine); noradrenergic and specific serotonergic antidepressant (mirtazapine); tricyclic antidepressants (doxepin, amitriptyline, trimipramine); non-benzodiazepine hypnotics (zopiclone); dopaminergic antiparkinsonian agents (levodopa); dopamine agonists (pramipexole); central nervous system stimulants (methylphenidate, lisdexamfetamine); norepinephrine–dopamine reuptake inhibitors (bupropion); and sedating antihistamines (promethazine) (PMID: 41503370, 11422727, 12532316, 8665541, 28346959, 104005, 9972389, 8599416, 171695).

Specific comment 11. Sleep-related breathing disorders (e.g., obstructive sleep apnea syndrome) should also be reported. If this data cannot be collected retroactively, please indicate this potential bias in the Limitations section.

Response: Sleep-related breathing disorders were not systematically assessed in this study as per the study protocol, and retrospective ascertainment would likely be strongly biased at the present time, given that for many participants, this would require recall such symptoms from several years ago. We have therefore added a statement in the Limitations section acknowledging that sleep-related breathing disorders, which may influence nightmare symptoms, could not be evaluated and may represent a potential source of bias. However, as for any other potential confounding factor discussed elsewhere in this response, randomization provides a protection against a systematic effect on treatment outcome. Please see page 8, line 299 (limitations): *"Sleep-disordered breathing was not systematically assessed in this study and therefore cannot be excluded as a contributor to nightmare symptoms, as obstructive sleep apnea has been associated with nightmares potentially through mechanisms such as intermittent hypoxia and REM-sleep fragmentation caused by recurrent respiratory events during sleep* (BaHammam et al. 2019, MaCall et al., 2022)".

Specific comment 12. You stated that seven SAEs occurred in the dronabinol group, which were "all assessed as unrelated." These included tachycardia, which led to a 2-day hospitalization. Tachycardia is a typical side effect of dronabinol and CBM. Why was tachycardia assessed as "unrelated" in this case?

Response: Tachycardia has been reported as a potential pharmacological effect of cannabinoids; the assessor and medical reviewer therefore carefully evaluated the possible relationship between this SAE and the study medication. In this case, some factors argued against a causal relationship. The study medication was not interrupted during or after the event, and the participant continued the same nightly dose of dronabinol without recurrence of tachycardia. Accordingly, no dechallenge or rechallenge pattern was observed. Furthermore, causality assessment was performed independently by the study physician and a medical reviewer in accordance with clinical trial safety procedures, and both concluded that the event was not related to the study medication. We have clarified this point in the revised manuscript on page 5, line 199 (Discussion): *"The tachycardia event was considered unrelated to study treatment, given the absence of recurrence despite continued treatment, in a patient without known cardiovascular disease."*

Specific comment 13. In the dronabinol group, no increase in cannabis withdrawal symptoms was observed at week 12 based on the total Marijuana Withdrawal Checklist (MWC). This is surprising insofar as a rebound phenomenon in REM sleep is known to occur after cannabis withdrawal, accompanied by an increase in the frequency and intensity of nightmares. I therefore recommend

paying particular attention to the “sleep disorders” item on the MWC checklist. Please analyse, report and discuss this point in more detail in the Discussion.

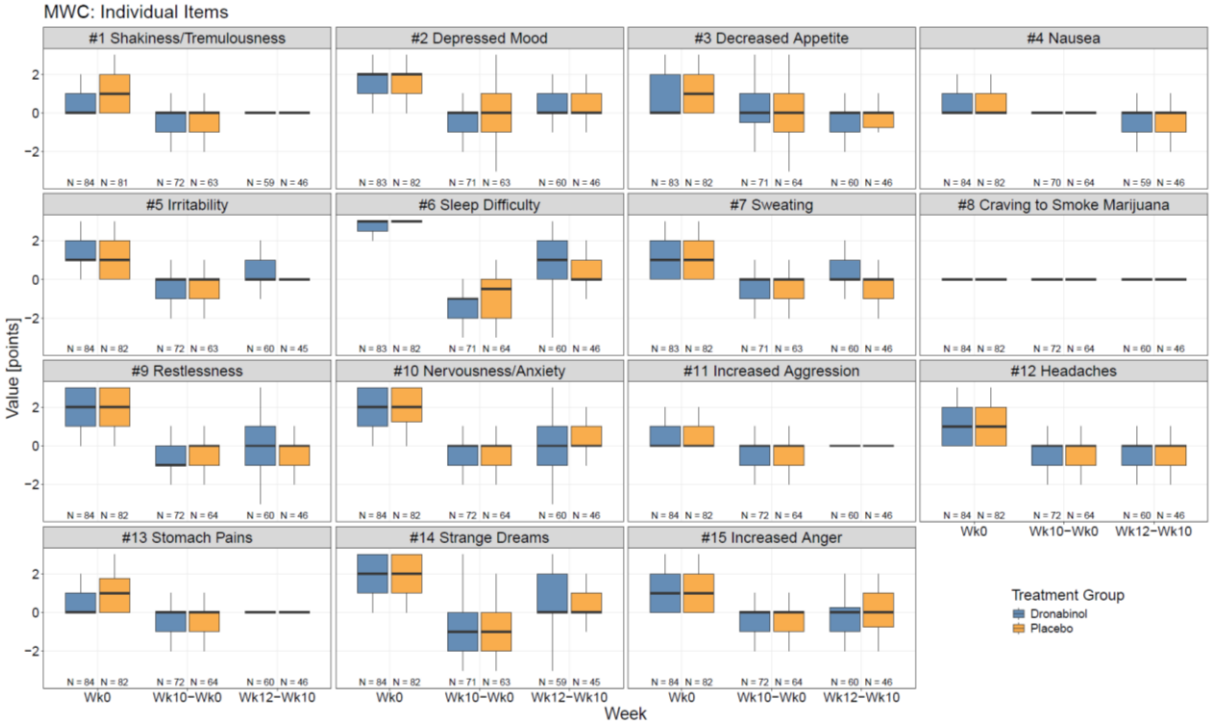
Response: To further examine potential withdrawal-related sleep effects, we have now analyzed the individual items of the Marijuana Withdrawal Checklist (MWC) and their time course between baseline, Week 10 (end of treatment), and Week 12 (follow-up) (see Supplementary Figure 6). Across most MWC items, distributions and median changes between Week 10 and Week 12 were similar in the dronabinol and placebo groups, with median changes of 0 for the majority of items. For the item “sleep difficulty,” we observed a slightly larger median increase in the dronabinol group (median change 1 vs. 0), although overall distributions were similar between groups and this small change did not translate into a meaningful difference in the total MWC score.

We also examined the item “strange dreams.” The median change from Week 10 to Week 12 was similar in both treatment groups, although the 75th percentile was somewhat higher in the dronabinol group, suggesting slightly greater variability in dream-related symptoms after treatment discontinuation. Importantly, other withdrawal-related symptoms such as irritability, anxiety, restlessness, and aggression did not show systematic increases after discontinuation of dronabinol. Taken together, the item-level analyses do not indicate a clinically relevant withdrawal pattern following cessation of dronabinol.

It should also be noted that the MWC assesses symptoms two weeks or more (for participants who discontinued study medication earlier but remained in the study) after cessation of study drug; therefore, a possible short-term REM rebound phenomenon immediately after discontinuation may not be fully captured by this assessment. REM rebound and increases in dream intensity after cannabis withdrawal have been described in previous studies (e.g., Velzeboer et al., 2025). Item-level time courses are provided in Supplementary Figure 6.

In addition, the following paragraph has been added to the Results (page 5, line 173): “*No increase in cannabis withdrawal symptoms on the Marijuana Withdrawal Checklist (MWC) was observed at week 10 or 12 in the dronabinol group (Supplementary Table 12, Supplementary Figs. 5, 6). However, a small increase in the MWC sleep-difficulty item occurred between weeks 10 and 12 (Supplementary Fig. 6)*.” and to the Discussion (page 7, line 275): *“Although REM rebound and increased dream intensity have been reported after cannabis withdrawal (Velzeboer et al., 2025), we observed only a small increase in the MWC sleep-difficulty item, which may in part reflect the timing of symptom assessment at two weeks or later after discontinuation.”*

Figure S6. Individual Items of the Marijuana Withdrawal Checklist (MWC).



Reviewer #2 (Remarks to the Author):

General comments: This manuscript was a multi-centre double blind placebo controlled trial of dronabinol for the treatment of nightmares in post-traumatic stress disorder. The primary outcome of the trial was significant, in that once daily dronabinol before bed for 10 weeks resulted in a significant reduction in the frequency and intensity of nightmares (CAPS IV B2) relative to placebo. There was also significant reduction in PGIC, while clinician assessed CAPS 5 overall scores and MADRS scores did not significantly differ between dronabinol and placebo. Importantly there were no serious adverse events. Overall, this is a well controlled and executed study and the findings are relevant and replicate and extend previous work from open label studies and underpowered RCTs. The comparative groups were appropriately balanced with respect to a host of criteria and the authors included, and recognized the importance, of a blinding assessment in the subjects given the psychoactive nature of the drug. I think this will be an important study for the field and will have significant influence on the focus of treatment of cannabinoid-based drugs for the management of specific PTSD symptoms, particularly sleep disturbances and nightmares. I only have two points which require addressing.

Response: We appreciate the reviewer's summary of the results and the quality of our trial and addressed the specific points raised as outlined below.

Specific comment 1: I cannot find the baseline values for the CAPS IV B2, CAPS V (overall PTSD score) and MADRS, the only data presented are difference scores. The baseline data need to be included here, at least in table, to ensure that the two populations didnt have significant baseline differences in symptom severity. If there were any differences, I couldnt find any discussion within the statistics section as to whether these baseline values were included as covariates to control for any influence that could have had on the outcomes.

Response: We agree that reporting baseline values is important to allow readers to assess potential differences between treatment groups. Baseline values of primary and secondary outcomes were presented in Table S4 in the Supplement. In response to this comment, we have now moved the baseline values for the primary and key secondary endpoints (CAPS-IV B2, CAPS-5 total score, and MADRS) to Table 1 in the revised manuscript. As shown in the table, there were no clinically relevant differences in baseline symptom severity between the treatment groups. The primary analysis was conducted using the prespecified mixed model for repeated measures (MMRM) based on change from baseline, with baseline values included as a covariate.

Specific comment 2: Perhaps this is a reflection of the nature of the question in the different versions, but why did the authors have the primary outcome measure be from CAPS IV but then the overall assessment of PTSD severity from CAPS 5? I couldnt find any mention of this within the methods section but would help to provide this information to understand the choice the outcome measures.

Response: We thank the reviewer for this important comment. The CAPS-IV B2 item ("recurrent distressing dreams") was selected as the primary endpoint to ensure comparability with prior pharmacological trials targeting PTSD-related nightmares. In particular, previous prazosin trials investigating nightmare-focused pharmacotherapy in PTSD consistently used this item as the primary outcome (e.g., Raskind et al. 2003, Raskind et al. 2007, Raskind et al. 2013, Raskind et al, 2018). Using the same item therefore allows the results of the present trial to be interpreted in the context of the existing literature.

The CAPS-IV B2 item assesses nightmare frequency and intensity on separate 0–4 scales, resulting in a combined score ranging from 0 to 8, whereas the corresponding CAPS-5 item provides a combined rating ranging from 0 to 4.

At the same time, overall PTSD severity was assessed using the CAPS-5 total score, reflecting the current DSM-5 diagnostic framework. CAPS-5 was therefore included as a key secondary outcome to capture global PTSD symptom severity according to current diagnostic standards, while CAPS-IV B2 provided a targeted measure of nightmare symptoms consistent with prior pharmacological trials.

To further address the reviewer's comment, we conducted an additional analysis using the corresponding CAPS-5 B2 item ("recurrent distressing dreams") within the same MMRM framework (new Supplementary Table 18). The results were consistent with the primary analysis and showed a statistically significant treatment effect in favor of dronabinol (least-squares mean difference −0.46; 95% CI −0.88 to −0.05; p = 0.028 in the full analysis set), with similar findings in sensitivity analyses. Because the study was prospectively designed around the CAPS-IV B2 item and the two instruments use different scoring ranges (0–8 vs. 0–4), the CAPS-5 B2 analysis should be interpreted as supportive.

We have clarified this rationale in the revised manuscript. In the Methods section, we now state (page 17, line 634): "*The CAPS-IV B2 item was selected to maintain comparability with prior pharmacological trials targeting PTSD-related nightmares (e.g., Raskind et al. 2018).*" In the Results section (post hoc), we added (page 6, line 218): "*In addition, results using the corresponding CAPS-5 B2 item were consistent with the primary analysis (Supplementary Table 18), supporting the robustness of findings across CAPS-IV– and CAPS-5–based assessments of nightmares.*"

Table S18: Supportive analysis using the CAPS-5 B2 item ("recurrent distressing dreams")

Sample	Outcome	Estimate	95% confidence interval		p-value
			CI-Lower	CI-Upper	
Complete Cases set	CAPS-5 B2 score from BL to wk 10	-0.54	-0.96	-0.11	0.014
Full Analysis Set	CAPS-5 B2 score from BL to wk 10	-0.46	-0.88	-0.05	0.028
Per Protocol Set	CAPS-5 B2 score from BL to wk 10	-0.57	-1.08	-0.07	0.026

Note: Results from MMRM analyses analogous to the primary analysis using the CAPS-IV B2 item. CAPS-5 B2 uses a scoring range of 0–4 compared with 0–8 for CAPS-IV B2.

Reviewer #3 (Remarks to the Author):

General comments: The authors find that dronabinol is more effective than placebo for the treatment of nightmares in individual with posttraumatic stress disorder (PTSD). The study builds upon prior, smaller, sometimes uncontrolled studies of cannabinoids (e.g., nabilone) for the treatment of PTSD-related nightmares or other PTSD symptoms. The size and rigor of this study are noteworthy advances over these prior works. The trial was pre-registered (NCT04448808) and the design and intended outcomes have remained true to the original submission. The writing is clear and cogent, and the manuscript is suitably succinct.

The finding that dronabinol was significantly superior to placebo on the primary outcome -- nightmare frequency and intensity -- is interesting and important. It is also collaterally supported by some of the secondary sleep-related outcomes, e.g., PSQI-A, where dronabinol was also superior to placebo. Of additional interest is the finding that overall improvement on the observer-rated CAPS-5 was not superior on dronabinol compared to placebo; but on the self-rated PCL-5, overall improvement was better on dronabinol compared to placebo. The authors may wish to discuss further this apparent discrepancy between the CAPS-5 and the PCL-5. Could it be that individuals with PTSD value improvement in sleep such that they feel they are overall improved, even if they do not experience broad-based improvement in PTSD symptoms?

Response: Differences between observer-rated CAPS-5 and self-reported PCL-5 outcomes have been described previously (Lee et al. 2022). Although both instruments assess DSM-5 PTSD symptom domains, the PCL-5 is a self-report measure capturing subjective symptom burden, whereas the CAPS-5 is a structured clinician-administered interview applying standardized probing, symptom attribution rules and severity thresholds (Lee et al. 2022). Prior work demonstrated that the two instruments are strongly correlated but not interchangeable and may yield different estimates of symptom severity or change over time (Lee et al. 2022). In addition, several psychometric studies reported that PCL-5 scores tend to be systematically higher than CAPS-5 scores and may therefore be more sensitive to subjective distress or broader affective symptom burden (Lee et al. 2024). We expanded the Discussion to address this potential explanation for the observed discrepancy by adding the following paragraphs to the revised manuscript. Page 6, line 238 (Discussion): "*Dronabinol did not significantly improve clinician-rated overall PTSD severity (CAPS-5) but was associated with improvements in self-reported PTSD symptoms (PCL-5). While the PCL-5 captures subjective symptom burden, the CAPS-5 applies structured probing and severity thresholds that may constrain score variability; the two instruments are strongly correlated but not interchangeable and can yield systematically different estimates (Lee et al. 2022, 2024).*"

Specific comment 1. If the CAPS-5 was used in the study, and the total score was used as one of the secondary outcomes, why was the primary (and some of the other secondary) outcome relying on the CAPS-IV B2 item? Why not the CAPS-5 B2 item?

Response: Thank you for raising this important point, which has also been brought up by Reviewer #2 (see detailed response to Rev#2, comment 2 above). In summary, the CAPS-IV B2 item ("recurrent distressing dreams") was selected as the primary endpoint to ensure comparability with prior pharmacological trials targeting PTSD-related nightmares. The CAPS-5 was used as a secondary endpoint to capture current PTSD diagnostic criteria. To address the reviewers' comments, we have now also added additional analyses using the CAPS-5 B2 item to the supplementary materials

(Supplementary Table 18) as well as paragraphs in the methods (page 17, line 634) and results (page 6, line 218) sections. Please see the detailed response above (Reviewer #2, specific comment 2).

Specific comment 2. Among the eight SAEs -- all of which occurred on drug -- two cases are referred to as "overdose of study drug". This clearly must be related to treatment, as it involves study drug, yet this and all the SAEs are reported as having been "not related to treatment". How is this possible? Was this an overdose with intent to harm oneself, overdose with intent to obtain additional effects of the drug, overdose in error? The nature and explanation of these two SAEs bears further explanation.

Response: We thank the reviewer for pointing this out. Seven SAEs occurred in the dronabinol group. The classification of an SAE as drug-related during our trial followed predefined standard criteria for causality assessments. In the present study, causality was assessed by the site investigator and independently reviewed by a sponsor-designated medical safety assessor. Two SAEs were coded as "study drug overdose". In the first case, a single episode of intentional excess intake occurred in a participant with comorbid borderline personality disorder and was clinically interpreted as a behavioral dysregulation event rather than a pharmacologically mediated toxicity. In the second case, the participant repeatedly took modest excess doses over several days without evidence of acute toxicity, dose-related worsening of symptoms, or a temporally linked pharmacological effect. In both cases, the events were therefore adjudicated by the responsible site physician and independently reviewed by a medical safety assessor as unrelated to the investigational product according to standard pharmacovigilance criteria. We have now added the following statements to the results part of the revised manuscript (page 5, line 191). *"Serious adverse events (SAEs) occurred in 7 patients (8.0%) receiving dronabinol and in none receiving placebo, all classified as not related to treatment by the investigators, according to standard causality-assessment procedures."* and (page 5, line 197): *"One case of dronabinol overdose reflected behavioral dysregulation rather than pharmacological toxicity; another showed no evidence of acute toxicity, dose-related worsening of symptoms, or a temporally linked pharmacological effect."*

Specific comment 3. There is one primary outcome and twenty secondary outcomes as listed with the clinical trial registry (NCT04448808). The authors do a good job of presenting all (or most) of these in the study. But it would be important for the authors to take into account this number of comparisons, and justify how they evaluate statistical significance in the context of 20 secondary outcomes. I don't necessarily need to see the authors apply a post-hoc correction, but I would like to see some rationale provided to the reader as to how to interpret and value this multiplicity of secondary outcomes.

Response: We agree that multiplicity needs to be considered when interpreting the key secondary outcomes. As prespecified in the Statistical Analysis Plan, a hierarchical testing strategy was applied. If the primary endpoint reached statistical significance at the alpha level adjusted for the interim analysis using an O'Brien–Fleming alpha spending function, the key secondary outcomes (CAPS-5, MADRS, and PGIC) were tested using the Hochberg procedure while maintaining a global significance level of 0.05. All remaining secondary endpoints were not included in the multiplicity-controlled testing procedure and are therefore interpreted descriptively and considered exploratory. No confirmatory claims are made for any of these outcomes, but we decided to present them in their entirety for full transparency. We have clarified this in the manuscript as follows (page 19, line 696, methods): *"Multiplicity was controlled using a prespecified hierarchical testing strategy: if the primary endpoint was significant, the three key secondary endpoints were tested simultaneously using the Hochberg procedure to control the family-wise type I error rate at a two-sided significance level of 0.05;* all remaining secondary endpoints were considered exploratory."

Reviewer #4 (Remarks to the Author):

General comment: The manuscript reports the results of a multicenter, double-blind, randomized, placebo-controlled trial of dronabinol (once daily before bedtime for 10 weeks) in adults with PTSD and nightmares. Overall, the results are clearly reported and largely adhere to established CONSORT and clinical trial reporting guidelines. The study addresses an important clinical question, and the statistical methods are generally appropriate. Below are several suggestions to improve clarity, transparency, and methodological rigor.

Response: We appreciate the reviewer's general comments on the trial's quality, rigor and reporting. We have addressed the suggestions to improve the manuscript as detailed below.

Specific comment 1. Toward the end of the "Trial Design and Oversight" subsection of the Methods section, please specify who had full access to the trial data, who performed the statistical analyses, and whether the authors had independent access to the data apart from the sponsor.

Response: We added to the subsection of the methods section (page 16, line 578): "*The trial was sponsored by Charité–Universitätsmedizin Berlin, Germany. Stefan Roepke, Robert Roehle, Erin D. Sprünken, and Stefanie Koglin had full access to all study data and are responsible for the integrity of the data and the accuracy of the analyses. Statistical analyses were performed by Robert Roehle and Erin D. Sprünken.*"

Specific comment 2. The trial used block randomization with a fixed block size of four. Please provide justification for using a fixed block size rather than varying block sizes, particularly given potential concerns about predictability in allocation.

Response: Randomization was performed using block randomization with a fixed block size of four to maintain balanced allocation between treatment groups throughout enrollment. The block length was concealed from the sponsor and all investigators across study sites until the end of the trial, thereby preventing predictability of treatment assignment and preserving allocation concealment. We have now added a statement to the methods section of the manuscript (page 17, line 606): *"The block size was concealed from the sponsor and investigators."*

Specific comment 3. The intention-to-treat (ITT) population was defined as all randomized patients who received at least one dose of study medication and had post-baseline data. Because this definition excludes randomized patients who did not receive treatment or did not provide post-baseline data, it does not meet the strict definition of ITT. Please either refer to this population as a modified ITT (mITT) population or justify why this definition is consistent with estimand and analytic framework of the trial.

Response: The primary analysis in our study was conducted in the full analysis set (FAS). We did not define a separate "intention-to-treat (ITT) population" in the manuscript; instead, the FAS was used as the principal analysis population following the intention-to-treat principle as described in the ICH E9 Statistical Principles for Clinical Trials guideline. According to ICH E9, the full analysis set should include all randomized patients whenever possible and should be "as close as possible to the intention-to-treat ideal," while allowing limited and justified exclusions, for example when a patient does not receive any study medication or when no post-randomization data are available. In our trial, two randomized patients met these predefined exclusion criteria: one patient did not receive study medication and one patient had no post-baseline assessment. These exclusions were prespecified in

the Statistical Analysis Plan. The resulting full analysis set therefore represents the closest practical approximation to the intention-to-treat principle. We have added to the methods section (see page 18, line 677): "*The primary analysis was conducted in the full analysis set (FAS), defined according to the intention-to-treat principle, corresponding to a modified intention-to-treat population, including all randomized participants who received at least one dose of study medication and had at least one post-baseline efficacy assessment.*"

Specific comment 4. The primary endpoint (change in CAPS-IV B2 score from baseline to week 10) was derived from a DSM-IV based instrument, whereas eligibility criteria were based on DSM-5. Please justify the choice of the DSM-IV B2 item as the primary outcome measure instead of using a DSM-5 based nightmare assessment and clarify the clinical rationale for this decision.

Response: Thank you for raising this important point, which has also been raised by Rev#2 (specific comment 2) and Rev#3 (specific comment #1). We have clarified the rationale in the manuscript and added additional analyses on the CAPS-5 B2 item. See above for details.

Specific comment 5. Missing data were handled using reference-based multiple imputation (jump-to-reference). Please clarify: (1) The underlying missing data assumption (e.g., MAR within treatment arm, treatment-policy estimand framework); (2) The rationale for choosing a reference-based (jump-to-reference) approach rather than standard MMRM alone; and (3) Whether the multiple imputation analysis constituted the primary analysis or a sensitivity analysis. Please also provide a brief explanation of the clinical and statistical rationale for this approach.

Response: Missing data for the primary endpoint were handled using reference-based multiple imputation with a jump-to-reference (J2R) strategy, as prespecified in the Statistical Analysis Plan. In this framework, patients who discontinue treatment are assumed to follow the outcome trajectory of the control (reference) group from the time of discontinuation onward. This approach reflects a clinically plausible Missing Not At Random (MNAR) scenario, as treatment discontinuation may occur due to perceived lack of efficacy. In addition, given the short duration of clinically relevant effects of oral dronabinol (Huestis 2007), with effects typically limited to several hours after administration, a rapid loss of treatment effect following discontinuation is expected, further supporting the assumption that patients subsequently follow the trajectory of the control group. In superiority trials, the J2R framework therefore represents a conservative assumption, as any potential treatment benefit is assumed to cease after treatment discontinuation. Each imputed dataset was analyzed using the prespecified mixed model for repeated measures (MMRM), and the resulting estimates were combined across imputations using Rubin's rule. While likelihood-based MMRM models can accommodate missing data under a Missing At Random (MAR) assumption, the reference-based imputation approach allows evaluation of treatment effects under the MNAR scenario described above. Thus, reference-based multiple imputation constituted the prespecified primary missing-data handling strategy within the primary analysis framework. Two hundred imputed datasets were generated for the analysis. A similar approach was applied for key secondary endpoints. We have clarified this description in the methods section, statistical analysis subsection, of the manuscript (see page 18, lines 687): "*Missing data for the primary endpoint were handled using prespecified reference-based multiple imputation (jump-to-reference), assuming that patients who discontinued treatment followed the control trajectory thereafter (a conservative Missing Not At Random (MNAR) assumption). Imputed datasets (n=200) were analyzed using MMRM and combined using Rubin's rule; this approach constituted the primary missing-data handling strategy and complements likelihood-based MMRM under a Missing At Random (MAR) assumption.*"

Specific comment 6. For the key secondary endpoints, the Hochberg procedure was used for multiplicity control. Please clarify whether this was implemented within a hierarchical testing

strategy contingent on significance of the primary endpoint. Also, explicitly describe the order of testing and the family of hypotheses controlled.

Response: Multiplicity control was prespecified in the Statistical Analysis Plan using a hierarchical testing strategy. First, the primary endpoint was tested at the significance level adjusted for the interim analysis using an O'Brien–Fleming alpha spending function. Only if the primary endpoint reached statistical significance, the key secondary endpoints were formally tested using the Hochberg procedure. The family of hypotheses for multiplicity control consisted of the three key secondary endpoints (CAPS-5, MADRS, and PGIC). These endpoints were considered of equal importance and were therefore tested simultaneously using the Hochberg procedure, controlling the family-wise type I error rate at a global significance level of 0.05. We have now clarified this testing strategy in the methods section of the manuscript (see page 19, line 696): *"Multiplicity was controlled using a prespecified hierarchical testing strategy: if the primary endpoint was significant, the three key secondary endpoints were tested simultaneously using the Hochberg procedure to control the family-wise type I error rate at a two-sided significance level of 0.05; all remaining secondary endpoints were considered exploratory."*

Specific comment 7. In the Results section, please clarify why the primary analysis population (n = 169) excludes two randomized patients (n = 171 total randomized). Although the CONSORT diagram provides information, a brief explanation in the Results text would improve clarity and transparency.

Response: We clarified in the results section (see page 3, line 112): *"From October 1, 2020 (first patient in) to January 2, 2025 (last patient out), 213 patients were screened at four centers; 171 underwent randomization, with 87 assigned to dronabinol and 84 to placebo. One patient assigned to placebo did not take the study drug and was excluded from the safety population (N = 170). The full analysis set comprised 169 patients: one patient did not take the study drug and one discontinued early, resulting in baseline-only data; both were assigned to placebo and were excluded from the primary analysis (Fig. 1)."*

Specific comment 8. The primary outcome was reported using Cohen's d. Please clarify whether this effect size was calculated using unadjusted means and pooled standard deviations or derived from the model-based (adjusted) estimates. If model-adjusted was used, please specify the covariates included in the model and list them in the corresponding table footnotes.

Response: Cohen's d was derived from the model-based treatment effect (LS means) estimated in the mixed model for repeated measures (MMRM), rather than from unadjusted group means. Specifically, the effect size was calculated using the adjusted treatment difference estimated by the MMRM, which included baseline CAPS-IV B2 score, visit, study center, treatment group, and the corresponding treatment-by-visit and baseline-by-visit interaction terms as fixed effects. This approach reflects the model-adjusted treatment effect and effect sizes were derived from the adjusted model estimates rather than from raw group means. We added to the statistical analysis subsection of the methods section (page 18, line 678): *"Cohen's d was derived from this model-based treatment effect and is presented with a two-sided 95% confidence interval."* Following the reviewer's suggestion, the covariates included in the model are now listed in the corresponding table 2 footnotes (see page 15, line 565). *"Analyses used a mixed model for repeated measures with baseline CAPS-IV B2 score, visit, study center and treatment group as fixed effects, including treatment-by-visit and baseline-by-visit interaction terms."*

Specific comment 9. Given the evidence of partial functional unblinding, it is unclear whether any sensitivity analyses were conducted adjusting for treatment-assignment guess. If such analyses were performed, please report them. If not, please discuss whether this may have influenced patient-reported outcomes.

Response: Because treatment assignment guess was assessed only after treatment completion and appears to reflect both perceived drug effects and clinical response, adjusting outcome analyses for treatment guess would introduce post-treatment bias and could obscure the causal interpretation of treatment effects. For this reason, we did not perform sensitivity analyses stratified by treatment assignment guess. We refer to this topic in the Discussion section, noting that *"…while underscoring the challenges of maintaining blinding in trials of psychoactive compounds and warranting cautious interpretation of patient-reported outcomes*" (page 7, line 290).

Specific comment 10. In Figure 2, please add the sample size at each visit for each treatment group. This would allow readers to better assess attrition over time and interpret longitudinal trends.

Response: Figure 2 presents model-based estimates derived from the prespecified mixed model for repeated measures (MMRM) analysis. Missing data were handled using a prespecified reference-based multiple imputation (jump-to-reference) approach prior to model estimation. Consequently, the trajectories shown in Figure 2 reflect model-based estimates from the imputed dataset rather than observed data at each visit, and visit-specific observed sample sizes are therefore not directly represented in this figure.

To enhance transparency regarding attrition and observed data, we provide Supplementary Figs. 2–4, which display the raw observed values for the primary and key secondary outcomes at each visit together with the number of participants contributing data at each time point.

Further, we added the following statement to the footnote of Fig. 2: "*Observed values for CAPS-IV B2, MADRS and CAPS-5 total scores, along with the number of participants contributing data at each time point, are shown in Supplementary Figs. 2–4."*

Responses to editorial comments:

- Please provide a paragraph in the Introduction highlighting the aims of the study and primary outcome(s) studied, before the Results section.

Response: We have added a dedicated paragraph to the end of the Introduction (page 3, line 108): “The aim of the THC-PTSD trial was to evaluate the efficacy and safety of dronabinol (THC, BX-1) in patients with PTSD-related nightmares. The primary outcome was the change from baseline to Week 10 in nightmare frequency and intensity, assessed using the Clinician-Administered PTSD Scale (CAPS-IV) B2 score (range, 0–8).”

- Please remove all references to Figures and Tables throughout the main text (e.g. ###Figure 1. CONSORT Chart###)

Response: All placeholder-style references have been removed.

- In the Subsection titled “Secondary Outcomes” in the Results section, please provide subheadings for the specific Secondary outcomes provided

Response: We have introduced structured subheadings within the “Secondary Outcomes” section to improve clarity and readability.

- The Sensitivity analyses section needs to specify what sensitivity analyses were done in the Main Text

Response: We have revised the Sensitivity Analyses section in the Main Text as described below to explicitly specify the sensitivity analyses performed.

Results (page 6, line 220): “Sensitivity analyses across the per-protocol set, under-medication set, complete-case set, last-observation-carried-forward approach, baseline-observation-carried-forward approach, and the originally planned ANCOVA model yielded consistent findings (Extended Data Table 4), supporting the robustness of the treatment effect of dronabinol on the primary endpoint.”

- You currently have 6 main display items, and 6 Supplementary figures. Please convert all of your supplementary figures to extended display figures to enhance visibility (each of which must fit on a single page). Please note that we can permit 6 main display items and 10 Extended Data display items.

Response: All Supplementary Figures have been converted to Extended Data Figures and reformatted to comply with the single-page display requirements. In addition, four Supplementary Tables have been converted to Extended Data Tables. The number of display items now complies with the journal limits.

- The supplementary material cannot include references. Please move the references in the supplementary material to the main reference list and ensure that this is appropriately cited in the main text.

Response: All references previously cited in the supplementary material have been moved to the main reference list and are now appropriately cited in the main text. Methods-only references are labeled accordingly.

- Please provide a specific section on Sample size calculations.

Response: A dedicated subsection describing sample size calculation has been added, including assumptions, effect size, power, and reference to the SAP. We added the following text to the Methods section under “Sample size calculation” (page 21, line 779): “Assuming a standardized mean difference (Cohen’s d) of 0.45 on the CAPS-IV B2 score at week 10, supported by a meta-analysis[68], a sample size of 158 participants (79 per group) would provide 80% power to detect a difference at a two-sided significance level of 5%. Allowing for 10% attrition, 176 participants were planned for enrollment. Sample size calculations were performed using nQuery version 8.2.1.0 and R including rpact version 4.2.0 (see also the statistical analysis plan in the Supplementary Information).”

- Please change the subheading “Participants” to “Eligibility criteria” in line 615 of Methods

Response: This has been corrected as requested.

- In the Methods, please provide information on how safety and adverse events were monitored throughout the trial, including if Data Safety Monitoring Board were involved.

Response: We have now added a new paragraph under the subheading “Safety and adverse events monitoring” describing the safety monitoring procedures, including adverse event collection, safety oversight, and the involvement of the Data Safety Monitoring Board (DSMB) in the interim analysis (page 21, line 763): “Participants were monitored throughout the study for adverse events, serious adverse events, vital signs, electrocardiograms (ECGs), laboratory parameters, urinary pregnancy tests, physical examinations, and cannabis withdrawal symptoms. Adverse events were collected and reported in accordance with Good Clinical Practice guidelines and categorized by investigators according to severity (mild, moderate, or severe) and relationship to study treatment. Treatment-emergent adverse events were defined as events occurring after randomization, from the first administration of study medication until the end-of-treatment visit at week 10. Safety monitoring also included assessment of cannabis withdrawal symptoms using the Marijuana Withdrawal Checklist at baseline and weeks 6, 10, and 12. Safety oversight included regular review of safety data by the sponsor and investigators and annual Development Safety Update Reports (DSURs), including risk assessments. An independent Data

Safety Monitoring Board (DSMB) was convened for the interim analysis and reviewed unblinded safety data at that time."

- In the Data Availability statement, simply stating that the data is not publicly available is not sufficient. Please provide individual participant data, data dictionaries, and any additional supporting documents needed to enable understanding, replication, or secondary analysis. If full public release is not possible due to ethical, legal, or institutional restrictions, please provide a clear explanation of these constraints. This should include which components of the dataset and documentation can be shared, the period during which the data will be accessible, the groups or individuals who may be eligible to apply for access, and the types of analyses for which access may be granted. Please also provide details on the mechanism for requesting access, including the application process, the criteria used to evaluate requests, and the means by which approved applicants will receive the data. can be obtained and any procedures for obtaining such data (e.g. formal approval,)

Response: The Data Availability Statement has been substantially expanded and now reads as follows (page 23, line 835):

Data Availability

Individual participant data (IPD), including de-identified datasets, data dictionaries, and relevant supporting documents, will be made available upon reasonable request for academic research purposes. Access will be granted following review and approval of a research proposal by the sponsor (Charité – Universitätsmedizin Berlin) and the principal investigator (Stefan Roepke), and is subject to a data use agreement in accordance with applicable data protection regulations.

Data will be available beginning 6 months after publication and for a period of 5 years. Requests should be directed to the first author. Proposals will be evaluated based on scientific merit, feasibility, and compliance with ethical and legal requirements. Approved applicants will receive access to the data via a secure data-sharing environment.

- In the Code Availability statement- if there are no custom codes generated for this study, please remove this section altogether.

Response: As no custom code was generated, this section has been removed.

- Please label the References list cited only in the Methods as "Methods-only References" list

Response: References cited exclusively in the Methods are now grouped under "Methods-only References."

Responses to Reviewers' Comments:

Reviewer #1 (Remarks to the Author):

I would like to thank the authors for their precise and transparent response to the comments on the first draft of their manuscript. The authors have provided a detailed response to my point 6 regarding the primary endpoint. To ensure that all readers have the opportunity to understand the authors' reasoning and rationale for deviating from the predefined plan to use ANCOVA and instead using MMRM, I would like to encourage the authors to explain this in more detail in the Statistics section. Please also indicate how the sample size was calculated according to the SAP. Furthermore, I would like to encourage the authors to include the comparison between MMRM and ANCOVA analyses in the appendix.

Response: We have added a more detailed description of this aspect to the Statistics section (page 22, line 795): "In the study protocol and previously published methods paper [54], the primary analysis was described using an ANCOVA model comparing CAPS-IV B2 scores at week 10 between treatment groups while adjusting for baseline CAPS-IV B2 score. In the Statistical Analysis Plan, the primary analysis approach was changed to a mixed model for repeated measures (MMRM). This modification was introduced to allow the use of all available post-baseline observations and to provide an efficient approach for the analysis of repeated-measures data in the presence of missing observations."

Further, the corresponding ANCOVA analyses were included in the table presenting the sensitivity analyses for the primary and key secondary outcomes (Extended Data Table 4). We added the following sentence to the Results section "Sensitivity analyses" (page 6, line 220): "Sensitivity analyses across the per-protocol set, under-medication set, complete-case set, last-observation-carried-forward approach, baseline-observation-carried-forward approach, and the originally planned ANCOVA model yielded consistent findings (Extended Data Table 4), supporting the robustness of the treatment effect of dronabinol on the primary endpoint."

Reviewer #2 (Remarks to the Author):

The authors have addressed the minor points I had raised and requested and the revisions in accordance with the other reviewers requests as well has improved the manuscript.

Response: We thank the reviewer for the positive evaluation and helpful feedback.

Reviewer #3 (Remarks to the Author):

The authors have done a remarkable job of responding to copious -- and important --

reviewer critiques and suggestions. The resultant revision is clear, concise, and of considerable interest.
I have no recommendations for changes, other than to suggest to the authors that randomization does not guarantee that confounders are equally distributed between groups, especially in a study with this modest (though large for PTSD) sample size. They may want to temper some of the statements that they make in several places. For example: "...randomization should have balanced potential variability related to food intake across treatment arms." Bottom line: randomization reduces the risk of imbalances between groups on measured (and unmeasured) confounders, but it is not a panacea.

Response: We agree and have revised the relevant statements to avoid implying that randomization guarantees balance between treatment groups. The revised wording now reflects that randomization reduces the likelihood of substantial imbalance in measured and unmeasured confounders, including potential variability related to food intake, but does not eliminate this possibility. We stated in the discussion section (page 8, line 306): “Although randomization reduces the likelihood of substantial imbalance in these factors between treatment groups, the trial does not allow conclusions regarding the extent to which prior exposure may have influenced treatment expectancy.” And later (page 8, line 319): “Although food intake can influence the absorption of orally administered cannabinoids [50], identical administration instructions and randomization reduced the risk of between-group differences related to food intake, while individualized dose titration may have further mitigated variability in exposure.”

Reviewer #4 (Remarks to the Author):

The revised manuscript addresses most of my previous concerns; however, three issues remain unresolved and require further attention:
Previous Concern #2: "The trial used block randomization with a fixed block size of four. Please provide justification for using a fixed block size rather than varying block sizes, particularly given potential concerns about predictability in allocation."
The revised manuscript notes only that the block size was 4 but does not provide explicit justification for choosing a fixed rather than variable block size. Authors should directly address the predictability risk associated with a fixed block size of 4, as this is a recognized methodological concern in randomized controlled trials. Please include a clear rationale for this design choice.

Response: We thank the reviewer for raising this important methodological point. We agree that fixed block sizes may increase the risk that blinded staff could predict treatment allocations for individual patients if the block size is known or can be inferred. In the present trial, however, the block size was not disclosed to investigators, raters, site personnel, or other staff involved in participant enrolment and assessment. The randomization schedule was generated centrally and securely stored. Treatment allocation remained concealed, and participants, investigators, and study staff remained blinded to treatment assignment until formal unblinding following database lock. The use of a fixed block size of four was chosen to maintain treatment balance within study centers while supporting efficient management, packaging, and shipment of blinded investigational medicinal product across centers. This was particularly relevant given the multicenter design, the possibility of variability in

recruitment per site and potentially low numbers of participants in some centers, and the limited shelf-life and logistical handling requirements of the investigational product. Consistent with ICH E9, the key safeguard against selection bias is the prevention of predictability of treatment assignment. ICH E9 states that investigators and other relevant staff should generally be blinded to the block length, and identifies randomly selected block lengths as one possible approach to achieve the same purpose. In our trial, allocation concealment was achieved by keeping the block size undisclosed and by using a centralized system for the generation and secure handling of the randomization schedule.

We have modified the Methods section of the manuscript as follows (page 20, line 701): “Participants were randomized at baseline in a 1:1 ratio using a centralized computer-generated block randomization with a fixed block size of four, stratified by study center. The block size was not disclosed to investigators, raters, site personnel, or other staff involved in participant enrolment or assessment. Allocation concealment was ensured through centralized randomization using SAS, version 9.4 (PROC PLAN), with sequences generated, securely stored, and held by an independent pharmacy; these were not accessible to the sponsor or investigators. Treatment allocation remained concealed, and participants, investigators, and study staff remained blinded to treatment assignment until formal unblinding following database lock. A fixed block size of four was selected to maintain treatment balance within centers while enabling efficient packaging, shipment, and use of blinded investigational medicinal product across centers, taking into account the multicenter design and logistical constraints related to the limited shelf-life and handling of the study medication. Identical packaging, labeling, appearance, and titration procedures were used for active drug and placebo.”

Previous Concern #9: "Given the evidence of partial functional unblinding, it is unclear whether any sensitivity analyses were conducted adjusting for treatment-assignment guess. If such analyses were performed, please report them. If not, please discuss whether this may have influenced patient-reported outcomes."
The revised manuscript still does not include a sensitivity analysis for the primary outcome adjusted for treatment-assignment guess. If such an analysis was not conducted, this should be explicitly acknowledged. Authors should also provide justification for why it was not performed and discuss the potential implications of this omission for the interpretation of patient-reported outcomes.

Response: We conducted additional sensitivity analyses to further examine the potential impact of treatment-assignment guess on the primary outcome. Prior to these analyses, we assessed the association between actual treatment assignment and treatment guess using Spearman’s correlation coefficient ($\rho = 0.349$), which did not indicate any substantial multicollinearity.

To explore the impact of treatment guesses, we first added treatment-assignment guess as a covariate to the primary MMRM model. This resulted in an improved model fit, suggesting that treatment guess accounted for part of the variance in outcomes.

Next, we performed an exploratory subgroup analysis analogous to the predefined subgroup analyses by including a treatment × time × treatment-guess interaction term, along with the corresponding lower-order interaction terms, using the multiple imputed datasets.

In this model, the estimated least-squares mean differences for the primary endpoint (dronabinol minus placebo) favored dronabinol in both subgroups: among participants who guessed placebo, the difference was −1.65 (95% CI, −3.18 to −0.13), and among those who guessed dronabinol, the difference was −1.04 (95% CI, −2.10 to 0.03). Importantly, the interaction between treatment guess and treatment effect was not statistically significant (difference between subgroups, 0.62; 95% CI, −1.22 to 2.46; p = 0.5063).

These findings suggest that although treatment-assignment guess may have influenced the magnitude of observed effects to some extent, the direction of effect consistently favored dronabinol across subgroups, with no evidence for a statistically significant differential treatment effect based on treatment guess. We therefore interpret these analyses as supporting the robustness of the primary findings.

We added the following sentence to results section of the manuscript (page 5, line 191): “A subgroup analysis according to treatment guess did not reveal a significant interaction effect (difference between subgroups: 0.62, 95% CI, −1.22 to 2.46; p = 0.51), while the estimated treatment effects within both subgroups consistently favored dronabinol (Supplementary Table 9).”

We added the following table to the appendix:

Table S9. Subgroup analysis by treatment-assignment guess

Contrast	Estimate	Lower 95% CI	Upper 95% CI	Standard Error	P value
Difference between treatment-guess subgroups [a]	0.62	-1.22	2.46	0.93	0.51
Participants guessing dronabinol: Dronabinol vs placebo	-1.04[b]	-2.10	0.03	0.54	
Participants guessing placebo: Dronabinol vs placebo	-1.65[b]	-3.18	-0.13	0.77	

[a] Corresponds to the treatment × time × treatment-guess interaction effect from the MMRM model.
[b] Negative values indicate lower CAPS-IV B2 scores in the dronabinol group relative to placebo.

Previous Concern #10: "In Figure 2, please add the sample size at each visit for each treatment group. This would allow readers to better assess attrition over time and interpret longitudinal trends."
This concern has not been addressed. Per-visit sample sizes for each treatment group must be included in Figure 2 to enable readers to appropriately assess attrition patterns and interpret longitudinal trends.

Response: Per-visit sample sizes for both treatment groups have now been added to Figure 2 (panels a, b, and c) to facilitate interpretation of attrition patterns and longitudinal trends.