Evaluation of Medical Cannabis and Prescription Opioid Taper Support for Reduction of Pain and Opioid Dose in Patients with Chronic, Non-Cancer Pain

NCT04827992

Version 3.4

First version: 12/20/2021

Latest version: 10/31/2025

Table of contents

Glossary

Abbreviation	Term
APA	American Psychological Association
CNCP	Chronic non-cancer pain
CUD	Cannabis use disorder
GEE	Generalized estimating equations
CB	Cannabis
MME	Morphine milligram equivalents
OUD	Opioid use disorder
PMP	Prescription monitoring program
POTS	Prescription opioid tapering support
WL	Waitlist

Design

The study will examine the efficacy of the addition of cannabis (CB) to prescription opioid tapering support (POTS) to help reduce pain and opioid use for patients with chronic non-cancer pain (CNCP). The study will contrast participants randomized to a 24-week period of either (1) receiving both POTS and CB post-baseline (**CB+POTS)** versus (2) receiving POTS post-baseline while wait-listed for receiving CB (**WL+POTS**).

The study aims to enroll up to 250 participants, adults aged 18 to 75 with CNCP endorsing >6 months of pain (neuropathic, nociceptive, or centralized pain) on stable prescription opioid doses of $\geq$ 25 MME/day for >90 days. Participants will be randomized in a 1:1 ratio at the therapy-group level (Therapy groups will consist of up to 6 participants, and all participants in a group will be randomized to the same condition to avoid cross-contamination). Therefore, it is expected that the CB+POTS and WL+POTS groups will each have up to 125 participants.

Analytic approach

Primary outcomes

Our co-primary outcomes will be…

1. The summed score (ranging from 0 to 30) of the 3-item Pain Enjoyment General Activity (PEG) scale (Krebs et al., 2009), where higher scores indicate greater pain severity and/or interference.
    - The PEG scores will be collected daily via self-report through a smartphone app from the baseline assessment to the end of the 24-week period (i.e., up to 168 observations per participant). All post-baseline daily observations for PEG scores through the week 24 visit will be analyzed.
2. Prescription monitoring program (PMP) verified opioid dose, in median daily morphine milligram equivalents (MME), defined over the "monthly" (30-day) interval preceding the final (week 24) study visit.
    - The observed median daily MME will be defined over each interval, as well as each interval between study visits, and 30 days after week 24 (totalling 8 intervals pre-baseline, between visits, and after week 24).
    - The observed median may be subject to artifacts in the PMP daily dose that arise commonly due to overlapping prescriptions or to filling prescriptions several days early or late. Therefore, an independent review will be conducted separately by two qualified clinicians who are blind to group assignment, to evaluate the 8 monthly intervals of PMP-reported daily MME in order to correct the MME for artifacts. These clinicians will independently verify the median daily prescribed MME for each month prior and following the baseline and week 24, resulting in 8 measurements. If the separate reviews by the clinicians are in consensus that a MME value other than the observed median dose should be used, the dose will be adjusted in analyses accordingly. If reviewers do not agree, they will discuss and try to reach a consensus. If no consensus can be reached, opioid dose will be defined as the observed median of the daily MME reported by the PMP over the specified 30-day intervals. We will compute inter-rater reliability scores for the clinical reviews and document all dose adjustments from the PMP dose.
    - We expect long periods of stable opioid doses over the 6-month trial. Therefore, our primary analysis will assess differences in prescribed dose over the 30-day interval preceding the visit at week 24 of the study (i.e., opioid dose over month 7 of the study), while accounting for dose in the interval prior to the baseline visit.
    - PMP-based opioid doses will serve as the dose for primary analysis. Self reported dose will be analyzed as a secondary outcome. It is hoped that the PMP dose will capture change in prescriptions and more stable long-term dose reduction than daily self reported dose.

Statistical model

We will analyze both outcomes using a linear regression model.

For the PEG score, coefficients and standard errors for the linear model will be obtained using a linear mixed effects model with random intercepts for each therapy group and subject to account for the nested hierarchical structure. If there are issues with model convergence due to the inclusion of two sets of random intercepts, then the models will instead be fit using only random intercepts for subject.

For the PMP verified opioid dose, a linear mixed effects model with random intercepts for each therapy group will be specified, and again if there are issues with model convergence due to the random intercepts then a model without random effects will be fit. The specification of covariates and confirmatory estimate of interest for each outcome is described below.

The p-value for each primary contrast will be computed via a t-test using the coefficient estimate and model-based standard error. The primary contrast testing for an effect of CB above and beyond POTS will be deemed statistically significant for $p < 0.025$, thereby ensuring an overall family-wise error rate of 5% despite two primary outcomes.

For each outcome, the model specification and key confirmatory effect of interest will be…

PEG scores: The model will include the following covariates: (a) a restricted cubic spline specification for change over days since baseline (with two spline knots placed at recommended data quantiles); (b) A participant's average PEG score over all measurements taken before and at the baseline visit; (c) A participant's median prescription opioid dose (MME) in the interval preceding the baseline visit; and the confirmatory effect of interest, (d) a dummy-coded contrast between WL+POTS (the referent, coded as 0) and CB+POTS (coded as 1), testing whether a constant effect of CB exists, averaged over all time points. In other words, we specify an additive model, adjusting for baseline levels and with main effects for a) the impact of CB and b) change over time, but no treatment by time interaction.

Opioid dose: The model will include the following covariates: (a) a participant's median prescription opioid dose (MME) in the interval preceding the baseline visit; (b) A participant's average PEG score over all measurements taken before and at the baseline visit; and the confirmatory effect of interest, (c) a dummy-coded contrast between WL+POTS (the referent, coded as 0) and CB+POTS (coded as 1), testing whether there is a significant difference in opioid dose at week 24 for WL+POTS versus CB+POTS, holding baseline dose constant (and adjusting for the other listed covariate(s)).

Missing data

We will address missingness using multiple imputation via chained equations (MICE), performed separately in each study group. However, participants who have fewer than 14 days (two weeks) of non-missing data will be excluded from the analysis (i.e., participants with less than 8.3% of the total number of possible observations will be excluded).

The following predictors will be used in all imputation models for all outcomes:

- A participant's age in years;
- A participant's biological sex (male versus female);
- A participant's median prescription opioid dose (MME) in the interval preceding the baseline visit;
- Number of baseline opioid use disorder (OUD) symptoms;
- A participant's average PEG score over all measurements taken before and at the baseline visit;
- Time specified as a restricted cubic spline

If additional variables are determined prior to data analysis to be predictive of missingness, they will also be included.

For PEG data, post-baseline daily scores will be imputed in a "long" format using a mixed effects imputation model.

For PMP opioid dose, the interval medians will be imputed as separate variables in a "wide" format.

Missing outcome data will be imputed 40 times, and analyses will be run using complete and imputed data for each imputation iteration, and results will be pooled according to Rubin's rule.

Intent-to-treat analysis

We understand that there may be some contamination between groups (e.g., some patients in the WL+POTS group may use CB, and some patients in the CB+POTS group may decide not to use CB). As this is a pragmatic trial, our primary analysis will be an intent-to-treat analysis, in which all participants will be analyzed by group (CB vs WL+POTS). This intent-to-treat analysis will *be representative of real-world, ecologically valid outcomes*, in which a clinician would recommend CB to a patient, and then the patient would come to a decision about whether CB was helpful, and act accordingly. Therefore, this type of analysis, designed for pragmatic trials such as this, will help inform real-world clinical decision-making. However, we do acknowledge that this intent-to-treat analysis cannot answer the question of whether CB has a biological effect on pain and/or opioid use.

In addition, for the week 24 PMP outcome we will conduct a per-protocol analysis to address the risk of bias by indication (e.g., patients in the WL+POTS group who are suffering worse pain may be more likely to use CB). We will compare those who used CB regularly (weekly or more) vs those who did not use (verified by negative

urine screens and no self-reported use) using a marginal structural modeling approach to correct for "confounding by indication" by weighting data by the inverse probability of being in the CB or non-user group.

Secondary analysis

As a secondary analysis, we will extend the additive model for PEG scores by fitting a model including a treatment by time interaction (i.e., the product of the contrast between CB+POTS and WL+POTS and the time trend). We will conduct an analysis of variance comparing the simpler additive model to the more complex interaction model. This test will be included in the Benjamini-Hochberg false discovery rate correction applied to the analyses of secondary outcomes, and if the associated Wald test is significant at $p < 0.05$, this will indicate the presence of a treatment by time interaction.

Sensitivity analyses

For the PMP outcome, we will perform a sensitivity analysis estimating the difference in mean of the monthly median opioid dose at week 24 using a linear mixed effects model fit to all post-baseline data points, with random intercepts for each therapy group and subject to account for the nested hierarchical structure. If there are issues with model convergence due to the inclusion of two sets of random intercepts, then the models will instead be fit using only random intercepts for subject. This model will be fit using the following covariates: (a) a restricted cubic spline specification for change over days since baseline; (b) A participant's average PEG score over all measurements taken before and at the baseline visit; (c) A participant's median prescription opioid dose (MME) in the interval preceding the baseline visit; (d) a dummy-coded contrast between WL+POTS (the referent, coded as 0) and CB+POTS (coded as 1), and (e) treatment by time interaction terms. From this model, the difference in adjusted mean of the monthly median opioid dose at week 24 between CB+POTS and WL+POTS will be estimated and a Wald hypothesis test will be performed.

In addition, for each primary outcome we will conduct a minimum of 2 sensitivity analyses.

1. We will examine if the direction and significance of the primary contrast between CB+POTS and WL+POTS is robust to the inclusion of additional covariates, specifically age (in years), biological sex (male versus female), number of baseline OUD symptoms.
2. We will examine if the direction and significance the primary contrast between CB+POTS and WL+POTS is sensitive to our treatment of missing data, by fitting the statistical model to the observed data only.

Note it may be necessary to include additional sensitivity analyses to address unanticipated developments during the course of the study.

Clinical significance

Examination of PEG scores and opioid does means that a combination of clinical outcomes is possible (see Table 1), which will indicate whether *CB is helpful* (e.g. decreases opioid doses and/or PEG scores), *CB is harmful* (e.g. increases opioid dose and/or PEG scores), or that *CB has no clear effect on opioid dose/PEG scores (no notable changes, or increases one outcome and decreases another)*. In the third condition, an exploratory analysis will evaluate costs/benefits of CB to the individual patient, measured via the proposed secondary outcomes.

Table 1: Decision table for each possible outcome

Decision	PEG scores at 6 months compared to Baseline	Opioid dose at 6 months compared to Baseline	Meaning
CB is beneficial	CB+POTS < WL+POTS	CB+POTS < WL+POTS	CB reduces PEG score AND decreases opioid dose
	CB+POTS < WL+POTS	ns	CB reduces PEG score and does not affect opioid dose
	ns	CB < WL+POTS	CB does not affect PEG score but decreases opioid dose
CB is harmful	CB+POTS > WL+POTS	CB+POTS > WL+POTS	CB increases PEG score and increases opioid dose
	CB+POTS > WL+POTS	ns	CB increases PEG score and does not affect opioid dose
	ns	CB+POTS > WL+POTS	CB does not affect PEG score and increases opioid dose
Individual costs/benefits should be evaluated	ns	ns	CB does not affect PEG score or opioid dose
	CB+POTS < WL+POTS	CB+POTS > WL+POTS	CB decreases PEG score but increases opioid dose
	CB+POTS > WL+POTS	CB+POTS < WL+POTS	CB increases PEG score but decreases opioid dose

Secondary Outcomes

Our secondary outcomes will be…

1. The summed score (ranging from 14 to 70) of the 14-item Quality of Life Enjoyment and Satisfaction Questionnaire – Short Form (Q-LES-Q-SF; Schechter, Endicott, & Nee, 2007), where lower scores indicate greater dissatisfaction with life.
2. The T-score (mean of 50 and SD of 10) of the 8-item Depression subscale of the PROMIS-29 (Cella et al., 2010), where higher scores indicate a greater degree of depression.

3. The T-score (mean of 50 and SD of 10) of the 7-item Anxiety subscale of the PROMIS-29 (Cella et al., 2010), where higher scores indicate a greater degree of anxiety.
4. The number of symptoms (ranging from 0 to 11) for Opioid Use Disorder (OUD), based on the DSM-5 Opioid Use Disorder Checklist (American Psychiatric Association [APA], 2013).
5. The number of symptoms (ranging from 0 to 11) for Cannabis Use Disorder (CUD), based on the DSM-5 Cannabis Use Disorder Checklist (APA, 2013).
6. Self-reported opioid dose in MME units collected daily via self-report through a smartphone app.

The secondary outcomes will be collected monthly during in-person study visits over the 24-week period (i.e., up to 7 observations per participant), except for CUD symptoms which are collected at weeks 0, 12, and 24, and self-reported opioid dose which is collected daily.

Statistical model

For each secondary outcome, we will use the same linear regression model, design matrix, and linear mixed effects modeling method as proposed for the primary outcome of PEG score with main effects specification for the treatment and each covariate, and a restricted cubic spline specification of time. Due to CUD symptoms only being measured at two post-baseline time points, time will be specified as a binary indicator (week 24 vs. week 12). In each model, the primary contrast testing for a constant effect of CB above and beyond POTS will be deemed statistically significant for $p < 0.05$ following an adjustment across all secondary outcomes using the Benjamini-Hochberg method, thereby ensuring a false-discovery rate of 5% despite multiple comparisons over nine secondary outcomes.

Missing data

Separately for each secondary outcome, we will use the same approach (multiple imputation via chained equations in a wide format) as specified for the PMP outcome, with the same set of additional covariates.

Sensitivity analyses

At a minimum, the 2 sensitivity analyses proposed for both primary outcomes will also be run for each secondary outcome. Again, note it may be necessary to include additional sensitivity analyses to address unanticipated developments during the course of the study.

Power

Our original power analysis stated that n =250 (100 participants completed in each group assuming 20% attrition) would be needed to achieve an 80% power to detect an 18% improvement in pain (PEG) scores and/or a 13% decrease in opioid dose.

- *PEG scores*: Power curve estimates were based on preliminary data, 3205 daily pain scores (a component of PEG scores) reported by 46 participants in the previous CB study over a period of 84 days (roughly 3 months). The mean (6.3) and standard deviation (3.1) for pain scores in the first two weeks was used to compute percent reduction. For 125 participants per group, we would have 80% power to detect a minimum percent reduction of 18% in PEG scores for the CB+POTS group above and beyond that for the WL+POTS group. Even with only 100 participants per group, we would have 80% power to detect a minimum percent reduction of 20% in PEG scores for the CB+POTS group above and beyond that for the WL+POTS group.
- *Opioid dose*: Power curve estimates were based on opioid dose data for the 145 PEG score patients extracted from Massachusetts General Hospital's 2017 records. We used the mean (88) and standard deviation (32) in morphine milligram equivalents (MME) for compute percent reduction. For 125 participants per group, we would have 80% power to detect a minimum percent reduction of 13% in opioid dose for the CB+POTS group above and beyond that for the WL+POTS group. Even with only 100 participants per group, we would have 80% power to detect a minimum percent reduction of 20% in opioid dose for the CB+POTS group above and beyond that for the WL+POTS group.

Given the slower-than-anticipated recruitment, we conducted a power re-estimation for this trial, and revised expectations to conclude that with n=120 (50 completed per group assuming 20% attrition), at 80% power, we will be able to detect a **28% improvement in PEG scores and/or a 20% decrease in opioid dose**, which are likely to be clinically significant. While final analyses will rely on linear regression and mixed effects to address correlated outcomes, because the key contrast of interest can be expressed as differences in mean between CB+POTS and WL+POTS, power can be approximated via standard methods for independent two-sample t-tests.

- *PEG scores*: Power curve estimates were based on preliminary data, 3205 daily pain scores (a component of PEG scores) reported by 46 participants in the previous CB study over a period of 84 days (roughly 3 months). The mean (6.3) and standard deviation (3.1) for pain scores in the first two weeks was used to compute percent reduction. For 60 participants per group, we would have 80% power to detect a minimum percent reduction of 25% in PEG scores for the CB+POTS group above and beyond that for the WL+POTS group. With only 50 participants per group, we would have 80% power to detect a minimum percent reduction of 28% in PEG scores for the CB+POTS group above and beyond that for the WL+POTS group. Research suggests that current chronic pain treatments can

result in an approximately 30% reduction in an individual's pain scores [1], and that a 30% reduction in pain can significantly improve function and quality of life [2].

- *Opioid dose*: Power curve estimates were based on opioid dose data for the 145 PEG score patients extracted from Massachusetts General Hospital's 2017 records. We used the mean (88) and standard deviation (32) in morphine milligram equivalents (MME) for compute percent reduction. For 60 participants per group, we would have 80% power to detect a minimum percent reduction of 19% in opioid dose for the CB+POTS group above and beyond that for the WL+POTS group. With only 50 participants per group, we would have 80% power to detect a minimum percent reduction of 20% in opioid dose for the CB+POTS group above and beyond that for the WL+POTS group.

Software

All analyses will be done using the statistical software R (version 4.1.1; R Core Team, 2021) and integrated development environment RStudio (version 2020.9.0.351; RStudio Team, 2021). Data will be prepared using the R packages ‘dplyr’ (version 1.0.7; Wickham, François, Henry, & Müller, 2021) and ‘tidyr’ (version 1.1.4; Wickham, 2021). Models will be fit using the R package ‘geepack’ (version 1.3-2; Højsgaard, Halekoh, & Yan, 2006). Missing data will be imputed using the R package ‘mice’ (version 3.13.0; Van Buuren & Groothuis-Oudshoorn, 2011). Reproducible code and de-identified data will be organized using the R package ‘targets’ (version 0.8.1; Landau, 2021) and Gitlab (version 14.6.7; Gitlab Team, 2022).

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). https://doi.org/10.1176/appi.books.9780890425596.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57* (1), 289-300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Cella, D., Riley, W., Stone, A., Rothrock, N., Reeve, B., Yount, S., ... & PROMIS Cooperative Group. (2010). The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *Journal of Clinical Epidemiology*, *63* (11), 1179-1194. https://doi.org/10.1016/j.jclinepi.2010.04.011.

Gitlab Team (2022). *Gitlab* (Version 14.6.7) [Computer software]. https://about.gitlab.com/

Hays, R. D., Spritzer, K. L., Schalet, B. D., & Cella, D. (2018). PROMIS®-29 v2. 0 profile physical and mental health summary scores. *Quality of life Research*, *27*(7), 1885-1891. https://doi.org/10.1007/s11136-018-1842-3.

Højsgaard, S., Halekoh, U., & Yan, J. (2006). The R package geepack for generalized estimating equations. *Journal of Statistical Software*, *15*, 1-11. https://doi.org/10.18637/jss.v015.i02.

Krebs, E. E., Lorenz, K. A., Bair, M. J., Damush, T. M., Wu, J., Sutherland, J. M., ... & Kroenke, K. (2009). Development and initial validation of the PEG, a three-item scale assessing pain intensity and interference. *Journal of General Internal Medicine*, *24* (6), 733-738. https://doi.org/10.1007/s11606-009-0981-1.

Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73* (1), 13-22. https://doi.org/10.1093/biomet/73.1.13.

Landau, W. M., (2021). The targets R package: A dynamic Make-like function-oriented pipeline toolkit for reproducibility and high-performance computing. *Journal of Open Source Software*, *6* (57), 2959. https://doi.org/10.21105/joss.02959

R Core Team (2021). R: A language and environment for statistical computing [Computer software]. https://www.R-project.org.

RStudio Team. (2021). *RStudio: Integrated development environment for R* (version 2020.9.0.351) [Computer software]. RStudio, PBC. http://www.rstudio.com/

Schechter, D., Endicott, J., & Nee, J. (2007). Quality of life of 'normal' controls: Association with lifetime history of mental illness. *Psychiatry Research, 152* (1), 45-54. https://doi.org/10.1016/j.psychres.2006.09.008.

Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*, 1-67. https://doi.org/10.18637/jss.v045.i03.

Wickham, H., François, R., Henry, L. & Müller, K. (2021). *dplyr: A grammar of data manipulation* (version 1.0.7). [Computer software] Retrieved from https://cran.r-project.org/web/packages/dplyr/index.html

Wickham, H. (2021). *tidyr: Tidy messy data* (version 1.1.4). [Computer software] Retrieved from https://CRAN.R-project.org/package=tidyr