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

First version: 12/20/2021

Latest version: 5/02/2022

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 will be analyzed.
2. Prescription monitoring program (PMP) verified opioid dose, in mean daily morphine milligram equivalents (MME).
    - We expect little variation in opioid dose until the conclusion of the treatment regime. Therefore, we will only analyze opioid doses reported during baseline and at week 24 of the study.
    - Note that if participants and their doctors decide to reduce dose at week 24, we will use the reduced dose even if the new dose cannot immediately be implemented (e.g., due to delays in scheduling and refilling prescriptions) to ensure accurate representation of change.

Statistical model

We will analyze both outcomes using a linear regression model. Coefficients and standard errors for the linear model will be obtained using generalized estimating equations (GEE; Liang & Zeger, 1986). Note the GEE approach provides robust standard errors and well-calibrated p-values (i.e., a family-wise error rate of 0.05) even when distributional assumptions are violated and when heteroscedasticity is present. We will assume data are clustered over participants, and that the observations for a participant (pooled over each month in the case of daily PEG scores), are uncorrelated (The GEE method is also robust to misspecification of the correlation structure for a participant's observations). The p-value for the primary contrast will be computed via a z-test using the mean estimate and a robust standard error computed via the sandwich estimator. The primary contrast testing for a constant 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 key confirmatory effect of interest will be…

PEG scores: 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. Additionally, we will include the following covariates: (a) A quadratic trend for change over days, consisting of a z-score for days since baseline (the linear component) along with the same z-score raised to the power of two (the quadratic component); (b) A participant's PEG score at the baseline

visit (converted to a z-score); (c) A participant's prescription opioid dose (MME) at the baseline visit (converted to a z-score). In other words, we assume a conservative 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 treatment (WL+POTS versus CB+POTS) by time (baseline versus week 24) interaction, testing whether there is a significant reduction in opioid dose at week 24 for CB+POTS above and beyond any reduction for WL+POTS. Main effects will be dummy-coded (WL+POTS coded as 0, CB+POTS coded as 1; baseline coded as 0, week 24 coded as 1), and the interaction will be defined as the product of the two. Additionally, we will include as a covariate a participant's PEG score at the baseline visit (converted to a z-score).

Missing data

The GEE method is robust to data missing completely at random (MCAR), but it is more likely that data will be missing at random (MAR). Therefore, we will address missingness using multiple imputation via chained equations (MICE). 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). All missing post-baseline outcome values will be imputed for opioid dose. However, for daily PEG scores, when outcome data is missing over multiple days in a row, the first and final day in the run will be imputed, with the remainder excluded (to reduce computational burden and ensure imputed values do not have excessive influence on analyses). Missing outcome data will be imputed 40 times, using, at a minimum, the following predictors:

- A participant's age in years;
- A participant's biological sex (male versus female);
- A participant's prescribed opioid dose (MME) at the baseline visit;
- Number of baseline opioid use disorder (OUD) symptoms;
- A participant's PEG score at the baseline visit;
- A participant's type of pain (neuropathic, nociceptive, or centralized pain);
- The outcome value on the previous entry (i.e., lag 1).

Continuous variables (except for the lag 1 term) will be converted to z-scores. Categorical variables will be first effects coded and then converted to z-scores. If additional variables are determined prior to data analysis to be predictive of missingness, they will also be included. 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.

Sensitivity analyses

We will conduct a minimum of 4 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, and pain type (neuropathic, nociceptive, or centralized pain). Categorical effects will first be effect-coded (-1 for the referent level, 1 for the specified level, and 0 otherwise) and then all covariates will be converted to z-scores.
2. We will test our assumption of an additive model for PEG scores by fitting a model that includes a treatment by time interaction (i.e., the product of the contrast between CB+POTS and WL+POTS and the two covariates for the quadratic time trend). We will conduct an analysis of variance comparing the simpler additive model to the more complex interaction model – if the associated Wald test is significant at $p < 0.05$ following a correction using the Benjamini-Hochberg method (Benjamini & Hochberg, 1995), this will indicate the presence of a treatment by time interaction.
3. We will examine if the direction and significance the primary contrast between CB+POTS and WL+POTS is robust to our treatment of missing data by fitting the statistical model to the observed data only.
4. We will conduct an as-treated 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 examine CB without regard to treatment group assignment, instead examining those who used CB regularly (weekly or more) vs those who did not use (verified by negative urine screens and no self-reported use). We will correct for “confounding by indication” by weighting data by the inverse probability of being in the CB or non-user group.

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 and then averaged over each month (opioid dose is not expected to vary substantially day to day).

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).

Statistical model

We will use the same linear regression model, design matrix, and GEE method as proposed for our primary outcomes. Specifically, we will use the same statistical model used with the PEG scores (note by necessity the linear and quadratic time trends will be defined over monthly visits). 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

We will use the same approach (multiple imputation via chained equations) as specified for the primary outcomes (specifically, the approach used with PEG scores).

Sensitivity analyses

At a minimum, the 4 sensitivity analyses proposed for the 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

While final analyses will rely linear regressions robust to clustering and heteroscedasticity, because the key contrast of interest is the mean difference between CB+POTS and WL+POTS, power can be approximated via standard methods for independent samples t-tests. The target sample size was 125 participants per group, or 100 participants under a worse-case scenario of 20% attrition. A power curve for each outcome was computed, plotting the required sample size for 80% power against the associated minimum detectable percent reduction in the outcome measure.

- *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.

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