Supplementary Online Content

Lintzeris N, Bhardwaj A, Mills L, et al; Agonist Replacement For Cannabis Dependence (ARCD) study group. Nabiximols for the treatment of cannabis dependence: a randomized clinical trial. *JAMA Intern Med.* Published online July 15, 2019. doi:10.1001/jamainternmed.2019.1993

This supplementary material has been provided by the authors to give readers additional information about their work.

© 2019 American Medical Association. All rights reserved.

eTable 1. Comparison of Characteristics of Participants Who Were Randomized but Received No Medication to Those Who Were Randomized and Commenced Treatment

	Randomised No Meds (n=9)	Randomised w/ Meds (n=128)
Variable	M (SD)	M (SD)
Age	26.6 (5.3)	35.0 (11.0)
Age of First Use	15.3 (1.7)	15.5 (3.9)
Age of First Regular Use	17.8 (2.7)	19.2 (6.3)
Duration Since First Regular Use (yrs)	15.7 (9.8)	9.5 (5.0)
ICD Cannabis Dependence Score (0-10)	8.6 (1.3)	8.7 (1.5)
	n (%)	n (%)
Gender	4 (44.4)	5(55.6)
Brief Psychiatric Rating Scale 18[a]		
Not ill (<31)	8 (100.00)	101 (84.2)
Mildly ill (31-40)	0 (0.00)	10 (8.3)
Moderately Ill (41-52)	0 (0.00)	2 (1.7)
Extremely Ill (≥53)	0 (0.00)	7 (5.8)

***a**: For BPRS thresholds, see Leucht et al., (2005); http://bjp.rcpsych.org/content/187/4/366. Percentages based on numbers excluding missing values: Randomised but no meds = 8 (1 missing), and Randomised but received meds = 120 (8 missing).*

© 2019 American Medical Association. All rights reserved.

eAppendix 1. Site Comparisons of Biohistorical Values at Baseline

Analysis: To test for pre-existing site differences in relevant biohistorical variables that might bias results, separate regressions, with study site as the sole predictor, were performed on outcomes: (1) participant age, (2) gender, (3) frequency of cannabis use prior to study commencement, (4) quantity of cannabis use prior to study commencement, (5) ICD-10 cannabis dependence score, (6) age of first cannabis use, and (7) duration of regular cannabis.

Results: There were no significant differences across sites in any of the important baseline cannabis dependence indicators except in duration of regular cannabis use ($P<0.001$), driven by the Newcastle site, with an average duration of use of 21.7 years, compared to 11.3 years, 15.6, and 15.7 years at the remaining three sites. Average participant age was also significantly higher at this site (40.6 vs 30.2, 34.9, and 36.0 years respectively).

© 2019 American Medical Association. All rights reserved.

eAppendix 2. Primary Analysis: Effect of Treatment on Total Days of Cannabis Use Across the 12-Week Trial

Below are the results of the sensitivity analysis, comparing the effect of Treatment on total days used across three different datasets, Modified Intention to Treat, Per-Protocol, and Multilevel Multiply Imputed. Note that although the ANCOVA performed on each of these datasets also included Site, Treatment x Site interaction, and Baseline Use in the model, coefficients for these effects will not be reported as: (1) they are not the primary effects of interests, (2) pooled *P*-values could not be calculated manually for the omnibus Site and Treatment x Site effects from the ANCOVA performed on the multiply imputed data (see below).

eTable 2. Effect of Treatment on Total Days Used From Three Different Datasets, Controlling for Site, Treatment x Site Interaction, and Baseline Use

	F	Estimated Difference	95% CI	*P*
Treatment				
ITT[a]	6.07	18.6	3.5, 33.7	**0.017**
Per Protocol[b]	5.75	20.3	3.2, 37.4	**0.021**
Multilevel MI[c]	3.92	10.6	1.0, 20.2	**0.039**

a: *The participants who attended all four research interviews (Placebo: n = 36; Nabiximols: n = 31).* ***b****: The participants who took medication for the entire study period and had four complete research interviews (Placebo: n = 25; Nabiximols: n = 26).* ***c:*** *Random slopes multilevel multiple imputation used (see Schafer, 1997; van Buuren, 2011). Based on twenty imputations. All 128 participants who started the trial were imputed (Placebo: n = 67; Nabiximols: n = 61).*

eAppendix 3. Note on Multiple-Imputation Analysis

Obtaining multiply-imputed values for total days used required multi-level (i.e longitudinal) imputation, with participant ID as the clustering factor and time as the random effect. This imputation was performed on a person-period (i.e. long-form) version of the dataset and resulted in twenty imputed datasets, where all 128 participants had imputed values of days used for all four time points (at 0 weeks, 4 weeks, 8 weeks and 12 weeks). The imputed long-form datasets were then converted back to wide form and a total days used value obtained for each participant by summing the values of days used for each participant across the four time points.

Because the imputation was performed on a dataset of a different form to the dataset used to create the total days used variable, software could not be used to calculate pooled statistics for the ANOVAs performed on each of the twenty imputed datasets. Thus these calculations had to be performed manually. Calculating pooled *F* statistics is relatively simple: the pooled *F* value for each effect in the ANOVA (Treatment, Site, Treatment x Site, and baseline use) is the average of the *F* values for the effect in question across all twenty datasets. Unfortunately calculating pooled *p*-values is not so simple, as it requires applying Rubin's rules (Rubin, 1987) to obtain the Barnard-Rubin adjusted degrees of freedom and then applying those degrees of freedom to the empirical *F* distribution. Calculating the Barnard Rubin adjusted degrees of freedom requires calculating the pooled standard error, something that is, at present, only possible for individual comparisons within a regression (which have standard error for each comparison) not for *omnibus* tests (which only have mean squared error for each omnibus effect). Calculating *P*-values is possible if the non-adjusted degrees of freedom are entered into the empirical *F*-distribution, however these *P*-values would be incorrect as they are not adjusted for the uncertainty captured by the Barnard-Rubin adjustment.

Fortunately, as mentioned above, in this study we were primarily interested in the effect of treatment when controlling for Site effects and Treatment x Site interactions. In this case the *P*-value for Treatment controlling for site is ***exactly*** the same in the *F*-test of the group main effect in the ANOVA as it is for the between-group comparison of Treatment in the analogous regression model, and the Barnard-Rubin adjusted degrees of freedom *can* be calculated for this comparison in the regression. Although individual pairwise comparisons of Site main effects and the Treatment x Site interaction *were* computed (and can be reported upon request), since they are not the primary effects of interest we have not reported them here.

© 2019 American Medical Association. All rights reserved.

eTable 3. Mean Days Used Across the 12-Week Trial by Site and Treatment Group

	Placebo			Nabiximols			Total
Site	*n* Week 0	*n* [a] Week 12	Days Used[b] M (SD)	*n* Week 0	*n* Week 12	Days Used M (SD)	Days Used M (SD)
Langton	24	14	66.9 (23.0)	20	10	43.0 (34.9)	57.0 (30.4)
St. George	13	8	45.8 (37.2)	10	6	35.0 (32.5)	41.1 (34.4)
Western Sydney	18	5	69.4 (25.1)	17	6	45.2 (32.9)	56.2 (30.9)
Newcastle	12	9	29.0 (34.1)	14	9	19.4 (28.3)	24.2 (30.8)
All Sites	67	36	53.1 (33.0)	61	31	35.0 (32.4)	44.7 (33.7)

***a**: Consisted of participants who provided valid frequency of use data at all four research interview: week 0 (baseline), week 4, week , and week 12.* ***b**: Out of a possible 84 (12 x 7 days)*

eAppendix 4. Interpretation

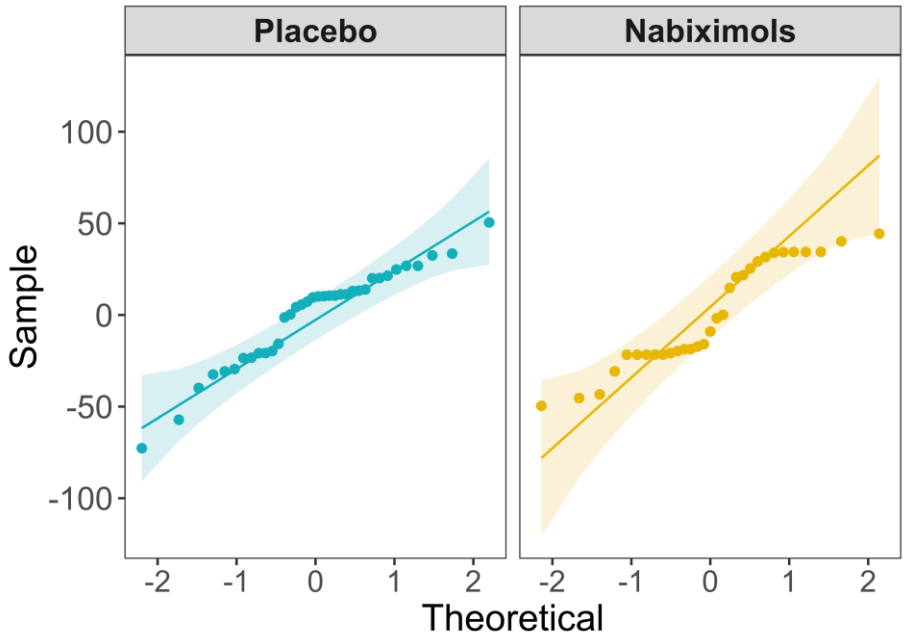
Main Effect of Treatment: The Placebo group used illicit cannabis an average of 53.1±33.0 days (out of 84 days) across the 12-week trial, compared to 35.0±32.4 in the Nabiximols group, a significant difference of 18.6 days after adjusting for baseline cannabis use (CI: 3.5–33.7, P=0.02).

Main Effect of Site: There was a main effect of Site (P=0.003), with participants at the Newcastle site using an estimated 33.6 fewer days over the 12-week trial, after adjusting for baseline use, than at the Langton (95%CI: 14.7–51.3; P<0.001) and 35.3 fewer days than at Western Sydney (95%CI: 11.2–56.8; P<0.001), with no significant differences between other sites.

Treatment x Site Interaction: The omnibus interaction between Treatment and Site was non-significant (P=0.94), suggesting that, despite site differences in total days used (and age, see eAppendix 1 above), the treatment effect – fewer total days used in the Nabiximols group compared to the Placebo group – was consistent across sites.

© 2019 American Medical Association. All rights reserved.

eFigure 1. QQ-Plots of Residuals for 12-Week Cannabis Use



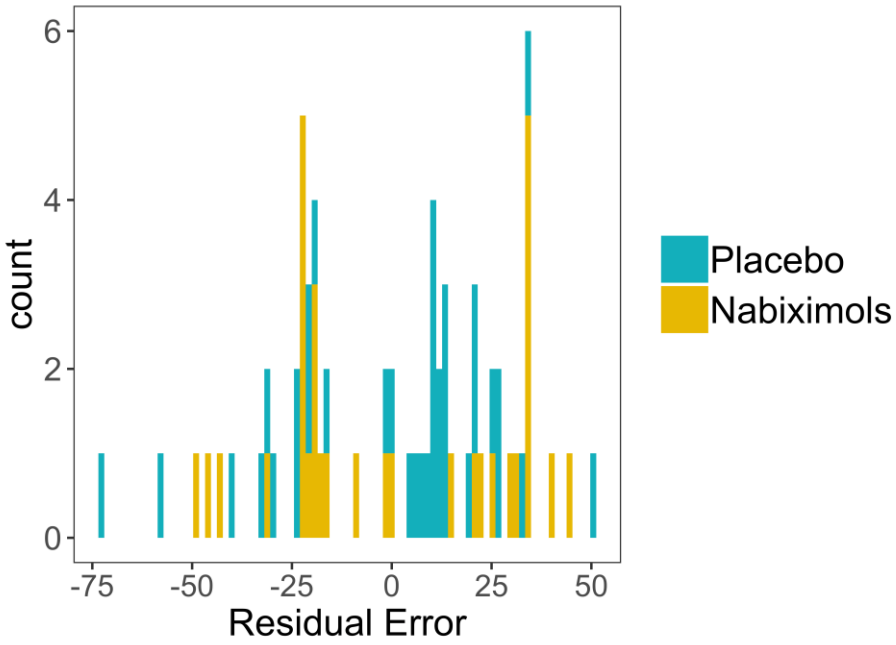
eAppendix 5. Analysis

These Q-Q plots are theoretical quantile-comparison plots for studentized residuals from the primary analysis, regressing 12-week cannabis use on (1) Treatment, (2) Site, (3) Treatment x Site, and (4) Number of Days Use in previous 4 weeks at Baseline. The plot has been faceted according to group. The straight diagonal comparison line is drawn on the plot through the quartiles of the two distributions. The closer to the line the observations fall the more normal the distribution is. The shaded area is the confidence envelope, computed by default by a parametric bootstrap. It is a rough guide to normality. The overwhelming majority of observations should lie inside the boundaries of this confidence envelope if the residuals are to be considered as being normally distributed, as is the case here.

© 2019 American Medical Association. All rights reserved.

eFigure 2. Histogram of Model Residuals for 12-Week Cannabis Use

Residuals appear near normal, with the majority in the centre of the range, however there appears to be another peak at approximately 27 on the x-axis to the right of the main peak in the centre of the distribution.



© 2019 American Medical Association. All rights reserved.

eTable 4. Adverse Event Reported by Group

	Placebo (n=67)	**Nabiximols** (n=61)	**Total** (*N*=128)
Number subjects reporting AE			
0	50 (74.6%)	46 (75.4%)	96 (75%)
1	10 (14.9%)	8 (13.1%)	18 (14.1%)
2	2 (3.0%)	2 (3.3%)	4 (3.1%)
≥3	5 (7.5%)	5 (8.2%)	10 (7.8%)
AEs reported (alphabetical order)			
Abdominal discomfort (pain, bloating)	3 (4.5%)	0	3 (2.3%)
Bad taste	0	2	2 (1.6%)
Chest pain	1 (1.5%)	1 (1.6%)	2 (1.6%)
Dizziness	1 (1.5%)	1 (1.6%)	2 (1.6%)
Dry mouth	1 (1.5%)	3 (4.9%)	4 (3.1%)
Flu-like illness	2 (3%)	2 (3.3%)	4 (3.1%)
Gastroenteritis	1 (1.5%)	1 (1.6%)	2 (1.6%)
Headache	2 (3%)	5 (8.2%)	7 (5.5%)
Mouth ulcer	2 (3%)	1 (1.6%)	3 (2.3%)
Oral paraesthesia / numbness	1 (1.5%)	2 (3.3%)	3 (2.3%)
Sleep problems	1 (1.5%)	1 (1.6%)	2 (1.6%)
Suicidal ideation	3 (4.5%)	0	3 (2.3%)
Tiredness / sedation	2 (3%)	3 (4.9%)	5 (3.9%)
Trauma or injury	1 (1.5%)	3 (4.9%)	4 (3.1%)
Vomiting	2 (3%)	2 (3.3%)	4 (3.1%)
Others (n=1 each)	7	7	14
Total	30	34	64
AE severity ratings[a]			
Mild	9 (28.2%)	18 (50%)	27 (39.7%)
Moderate	21 (65.6%)	14 (38.9%)	35 (51.5%)
Severe	2 (6.2%)	4 (11.1%)	6 (10.2%)
BPRS[b]	M (SD)	M (SD)	M (SD)
Baseline (*n*=120; Pl=63, N=57)[c]	23.6 (11.5)	24.1 (12.0)	23.8 (11.7)
Week 4 (*n*=79; Pl=42, N=37)	24.3 (13.6)	24.8 (11.2)	24.6 (12.5)
Week 8 (*n*=61; Pl=32, N=29)	20.8 (7.7)	21.8 (16.1)	21.3 (12.3)
Week 12 (*n*=42; Pl=22, N=20)	24.8 (7.8)	22.5 (11.1)	23.7 (9.5)

a: *Mild = awareness of AE event, but easily tolerated; Moderate = discomfort enough to cause interference with usual activity and may warrant intervention; Severe = incapacitating with inability to do usual activities or significantly affects clinical status, and warrants intervention.* ***b:*** *Brief Psychiatric Rating Scale.* ***c:*** *'Pl' = numbers in placebo group, 'N' = numbers in Nabiximols group.*

© 2019 American Medical Association. All rights reserved.

eTable 5. Comparison of Numbers of Each Type of Adverse Event and Total Adverse Events

	Mild	**Moderate**	**Severe**	**Extreme**	**Total**
Group	n(%)	n(%)	n(%)	n(%)	N
Nabiximols (*n* = 61)	18 (50.0)	14 (38.9)	4 (11.1)	0 (0.0)	36
Placebo (*n* = 67)	9 (28.2)	21 (65.6)	1 (3.1)	1 (3.1)	32
Total (*N* = 128)	27 (39.7)	35 (51.5)	5 (7.4)	1 (2.8)	68

Mild (awareness but easily tolerated) = 27; Moderate (causing interference with usual activity & may warrant intervention) = 35; Severe (incapacitating, inability to do usual activities or significantly affects clinical status) = 5 (excludes the SAE noted above). These AEs include depression (1, unrelated, Nabiximols), nightmares (1, possible, Nabiximols), suicidal ideation (1, possible, Placebo), vomiting (1, possible, Nabiximols) and headache (1, possible, Nabiximols)

eTable 6. Group Difference in Incidence Rate of AEs

	*IRR** (95% CI)	*P*
Adverse Events (Pl=67, N=61)	1.13 (0.50-2.57)	0.766

eAppendix 6. Analysis and Interpretation

A count, indicating total number of AEs of any type over the 12-week trial period, was calculated for each participant. Dispersion of these counts was assessed by comparing each group's mean to its standard deviation. Standard deviations were more than double group means in both groups. This combined with a dispersion parameter of $\alpha = 3.61$, indicated overdispersion, hence a negative binomial regression was used. Vuong's non-nested test (Vuong, 1989) indicated that the zero-inflated negative binomial regression offered no significant benefit over the non-zero-inflated model, therefore the non-zero-inflated model was used. Results of this analysis are presented in eTable 6. The incidence rate of aberrant behaviours was an estimated 13% higher in the Nabiximols group than the Placebo group, however this difference was not significant ($P = 0.766$).

© 2019 American Medical Association. All rights reserved.

eTable 7. Incidence Rates of Aberrant Behaviours by Group (ORBIT)

Count	Placebo (n=37) n (%)	Nabiximols (n=33) n (%)	Total (N=70) n (%)
0	23 (62.16)	26 (78.79)	49 (70.00)
1	9 (24.32)	5 (15.15)	14 (20.00)
2	5 (13.51)	1 (3.03)	6 (8.57)
3	0 (0.00)	1 (3.03)	1 (1.43)

eAppendix 7. Analysis

Aberrant behaviours were measured by the ORBIT questionnaire at the Week 12 research interview. A count, indicating number of aberrant behaviours over the 12-week trial period, was obtained for each participant. Dispersion of these scores was assessed by comparing each group's mean to its standard deviation. A dispersion parameter of $\alpha = 0.53$ indicated slight underdispersion. Test comparison using Vuong's non-nested test (Vuong, 1989) indicated that none of the more complex models (e.g. negative binomial or zero-inflated Poisson) outperformed the standard Poisson, therefore it was used to estimate between-group differences in rate of aberrant behaviours. Results of this analysis are presented in eTable 8.

eTable 8. Comparison of Incidence Rate of Aberrant Behaviours Between Groups

	IRR (95% CI)	*P*
ORBIT	0.59 (0.27-1.26)	0.176

eAppendix 8. Interpretation

The incidence rate of aberrant behaviours was an estimated 41% lower in the Nabiximols group than the Placebo group, but this difference was not significant ($P = 0.176$).

© 2019 American Medical Association. All rights reserved.

eTable 9. Comparison of Incidence Rate of Aberrant Behaviours Across Groups

ORBIT questionnaire responses	Placebo (n=37) n (%)	Nabiximols (n=33) n (%)	TOTAL (n=70) n (%)
1. I have asked my doctor for an early renewal of my prescription as I ran out early	1 (2.70)	1 (3.03)	2 (2.86)
2. I have used another person's cannabis medication, for example a friend or family members or bought it off the street	0	0	0
3. I have saved up my cannabis medication just in case I needed it later	2 (5.41)	1 (3.03)	3 (4.29)
4. I have tried to go to a different doctor to get more cannabis medication and did not tell my normal doctor	0	1 (3.03)	1 (1.43)
5. I have asked my doctor for another prescription because I had either lost my prescription, medication or had it stolen, or someone else had used it	0	0	0
6. I have given or sold my prescription medication to someone else	10 (27.03)	2 (6.06)	12 (17.14)
7. I have altered my dose in some other way, when I was not advised to do so by a health professional	6 (16.22)	4 (12.12)	10 (14.29)
8. I have taken my cannabis medication by a different route than was prescribed	0	1 (3.03)	1 (1.43)
Total	14 (37.84)	7 (21.21)	21 (30.00)

eAppendix 9. Analysis

A chi-squared test of independence was performed, testing for presence of a between-group difference in proportion of individual who engaged in *at least one* aberrant behaviour during the 12-week trial. Results of this analysis are presented in eTable 10.

eTable 10. Pearson's Chi-Square Test of Between-Group Difference in Rate of Aberrant Behaviours

	χ^2	*d.f*	*P*
ORBIT	1.57	1	0.210

eAppendix 10. Interpretation

Thirty percent (21/70) of respondents reported any aberrant medication behaviours over the medication period using the modified ORBIT questionnaire at Week 12 research interview. The Placebo group reported a higher rate of aberrant behaviours (37.84% c/w 21.21% Nabiximols), though a Pearson's chi-square test with continuity correction indicated the between-group difference in rate was not significant ($\chi^2_{(1)} = 1.57$; $P = 0.210$). The most common behaviours being 'giving or selling medication to someone else' (n=12, 17.1%) and 'altering my dose in some other way' (n=10, 14.3%).

© 2019 American Medical Association. All rights reserved.

eAppendix 11. Analysis 1: Testing the Ability of Self-Reported Days Use to Predict Urinary Cannabinoid Concentration

Urine samples taken at baseline, week 4, week 8, and week 12 were analysed for (−)-*trans*-Δ^9-tetrahydrocannabinol (THC), 11-Hydroxy-Δ^9-tetrahydrocannabinol (11-OH-THC), and 11-Nor-9-carboxy- 9-tetrahydrocannnabinol (THC-COOH). Each cannabinoid was adjusted according to creatinine concentration at the time of measurement using the procedure outlined in Baker and colleagues (2018). Analysis of urine drug tests results was conducted to examine the validity of self-reported illicit cannabis use at baseline and week 4,8 and 12 research interviews, and was therefore restricted to the Placebo group (the prescribed THC in the Nabiximols group prevented meaningful interpretation of urinary THC or metabolites). Two approaches were taken to the analysis of urinary cannabinoids.

Method 1: This method was a more general test; of whether self-reported days use predicted levels of the three metabolites. Creatinine adjusted urinary THC, 11-OH THC and THC-COOH levels from urine samples from all observations time points from all Placebo participants were compared against self-reported days illicit cannabis use collected on the same day at 4-weekly research interviews.. Due to the very large range and heavy positive skew in the data, creatinine-adjusted cannabinoid concentrations were Winsorized and then log-transformed prior to analysis. The log-levels of the three cannabis metabolites were then regressed on self-reported days use of illicit cannabis in the previous 28 days, with participant age and gender entered as covariates. These regressions were performed on data from each of the four measurement points in isolation (baseline, 4 weeks, 8 weeks, and 12 weeks). Results of this analysis are presented in eTable 11 and eFigure 3.

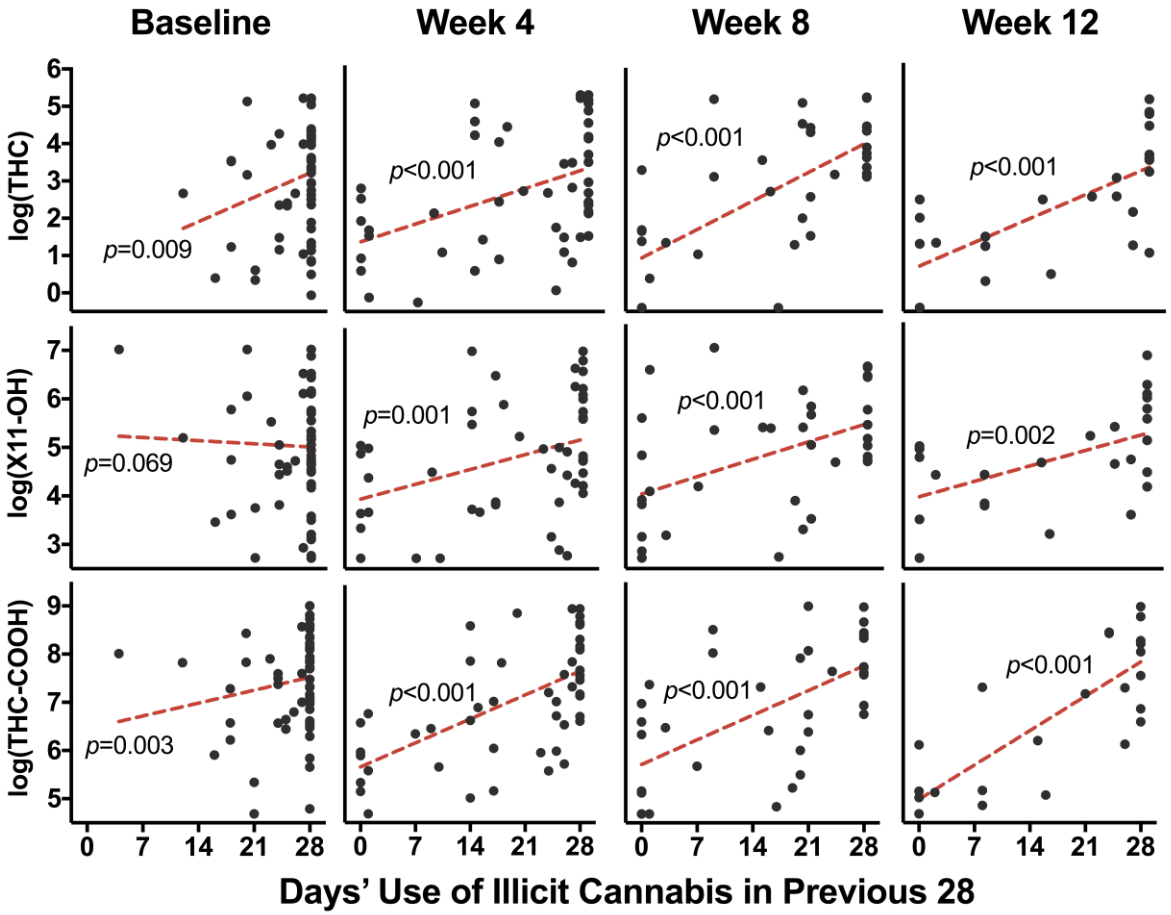
eTable 11. Results of Regressing Three Cannabinoids on Self-Reported Days Use of Illicit Cannabis, Age, and Gender at Four Different Timepoints in the Placebo Group

Data	n	Placebo M (SD)	Estimate[a]	*t*	*P*[b]	95% CI[a]
THC						
Baseline	63	57.1 (87.7)	1.13	6.92	*0.009*	1.03, 1.24
Week 4	51	55.4 (103.0)	1.08	16.97	***<0.001***	1.04, 1.12
Week 8	38	43.4 (54.4)	1.11	27.46	***<0.001***	1.07, 1.15
Week 12	28	26.8 (39.6)	1.11	19.27	***<0.001***	1.06, 1.17
11-OH-THC						
Baseline	63	331 (446)	1.07	3.35	0.069	0.99, 1.15
Week 4	51	305 (600)	1.05	11.90	**0.001**	1.02, 1.08
Week 8	38	257 (354)	1.06	14.44	***<0.001***	1.03, 1.09
Week 12	28	176 (195)	1.06	9.67	**0.002**	1.02, 1.10
THC-COOH						
Baseline	63	3302 (5289)	1.10	9.00	**0.003**	1.03, 1.18
Week 4	51	2711 (3773)	1.08	34.00	***<0.001***	1.05, 1.11
Week 8	38	2043 (2697)	1.08	27.67	***<0.001***	1.05, 1.11
Week 12	28	2119 (2773)	1.12	42.51	***<0.001***	1.08, 1.16

a: *Estimates are slope coefficients from regressing log-transformed values for each the cannabinoids on self-reported days use. Slope coefficients have been anti-logged to yield more readily interpretable estimates. Each coefficient is the estimated proportion increase in the outcome predicted from a 1-unit increase in the predictor when controlling for the other predictors. For example a coefficient of 1.10 means that an increase in self-reported use of one day predicts an estimated 10% increase in the creatinine-adjusted value of the cannabinoid in question.* ***b****: P < 0.05 in italics, P < 0.01 in bold, P <0.001 in bold italics.*

© 2019 American Medical Association. All rights reserved.

eFigure 3. Results of Regressing Three Cannabinoids on Self-Reported Days Use of Illicit Cannabis, Age, Gender, and Age First Used Cannabis, at Four Different Timepoints in the Placebo Group



eAppendix 12. Interpretation

180 observations were recorded across the four time points. When adjusting for age and gender, self-reported days use in the previous 28 days was a strong predictor of level of creatinine-adjusted THC, 11-OH-THC, and THC-COOH. Only in 11-OH-THC at baseline did self-reported days use fail to significantly predict cannabinoid levels. For the remaining analyses an increase of one day in self-reported use predicted an estimated increase in urinary cannabinoid levels of 5-13%.

© 2019 American Medical Association. All rights reserved.

eAppendix 13. Method 2

This approach attempted to verify self-reported cannabis use on a case-by-case basis, via the method described by Baker and colleagues (2018) for assigning recent abstinence from illicit cannabis use based on quantitative creatinine adjusted THC-COOH levels. Change scores were calculated for the THC-COOH variable by dividing Winsorized THC-COOH levels at each timepoint by Winsorized THC-COOH levels for the same participant at the previous timepoint (hence there were no change scores for baseline observations, which were not included in analysis). A binary variable was then calculated based on these change scores. Observations were recorded as abstinence from recent use (negative or '0') if THC-COOH levels dropped by more than 75% from the previous observation *or* if THC-COOH observations fell below 200 ng/ml. Any observations that did not meet this criterion were recorded as a positive ('1'), signifying recent illicit cannabis use.

It is worth noting that although Baker and colleagues also considered 50 ng/ml and 100 ng/ml as cutoff criteria, if we used either of these criteria in combination with the 75% reduction criteria no observations met criteria for abstinence. Although it is possible that every person in the placebo group who reported abstinence was not telling the truth, we favoured the more charitable interpretation—that this group were a clinical population of very heavy, long-term cannabis users whose baseline THC-COOH levels were persistently higher than Baker and colleagues' sample (who were not a clinical population). Thus we chose the higher threshold of 200 ng/ml as the absolute cut-off value to combine with the %-change score in assessing abstinence vs use.

A binary cannabis use variable was also calculated, where any number of days use was recorded as a positive ('1') and zero days use a negative (0). A contingency table was then calculated using these two binary variables (see left side of eTable 11). These analyses were concerned with verifying use or abstinence at the *observation level.* We were not concerned with *trajectory* of either cannabis use or THC-COOH levels, therefore time was not included as a factor in these models. Results of these two analyses are presented in eTables 12 and 13.

***ROC Analysis*:** ROC analysis tests the results obtained from one test against another test that is considered, for the purposes of analysis, to be the 'truth'. Neither self-reported abstinence or use *or* urinary THC-COOH levels could be considered reliable enough to confidently designate it the 'standard of truth' necessary for ROC curve analysis. Thus two versions of the analysis were conducted, one where the THC-COOH thresholds were set as the standard of truth, the other where self-reported days use were the standard of truth. The results of these two ROC analyses—sensitivity, specificity, percentage correctly classified, positive- and negative likelihood ratios, and area under the curve statistics—are presented on the right side of eTable 12.

eTable 12. Confusion Matrix of Numbers in the Placebo Group With Self-Reported Cannabis Use or Abstinence Versus Numbers Who Met Threshold Urinary THC-COOH Levels (Left) and Receiver Operating Characteristics Analysis (Right)

		THC-COOH[b]			If THC-COOH Thresholds are Treated as Truth					
		0	1		TPR[c]	TNR[d]	Correctly Classified	PLR[e]	NLR[f]	AUC[g]
Self-Reported Cannabis Use[a]	0	10	10	20	89%	50%	82%	1.78	0.22	0.69
	1	10	79	89	**If Self-Report is Treated as Truth**					
		20	89	109	89%	50%	82%	1.78	0.22	0.69

***a**: 0 = Zero self-reported cannabis use days in previous 28 days; 1 = one day or more of self-reported illicit cannabis use in previous 28 days. **b**: 0 = THC-COOH levels <25% of levels at previous measurement or < 200 ng/ml; 1 = THC-COOH levels ≥25% of levels at previous measurement and ≥ 200 ng/ml. **c**: TPR = True Positive Rate = Sensitivity. **d**:TNR = True Negative Rate = Specificity. **e**: PLR = Positive Likelihood Ratio = True Positive Rate/False Positive Rate. **f**: NLR = Negative Likelihood Ratio = False Negative Rate/True Negative Rate. **g**: AUC = Area Under the Curve.*

***Logistic Regression*:** The binary THC-COOH variable was regressed on the binary self-reported cannabis use variable, with age and gender included as covariates. Results of this analysis are presented in eTable 13.

eTable 13. Relationship Between Self-Reported Cannabis Use and Urinary THC-COOH Levels

	*OR** (95% CI)	*P*
Abstinence	7.9 (2.5, 26.0)	<0.001

**Odds ratio of testing negative for THC-COOH after reporting abstinence in the previous 28 days.*

eAppendix 14. Interpretation

In the Placebo group there were 109 observations with current and previous valid measurements of both self-reported use and urinary THC-COOH levels. In 82% of those observations self-report matched the THC-COOH threshold results. When using the criteria for a negative result of a ≥75% reduction or absolute levels of THC-COOH less than 200 ng/ml, the odds of participants testing negative for THC-COOH were eight times higher if they reported no cannabis use in the previous 28 days than if they reported any use (OR = 7.9; 95% CI: 2.5, 26.0; $P < 0.001$). The odds of participants testing positive after reporting any use were identical to those of testing negative after reporting

© 2019 American Medical Association. All rights reserved.

abstinence. Almost identical results were also obtained when outcome and predictor were reversed, and self-reported cannabis use was regressed on the binary THC variable—with the odds of reporting abstinence also approximately 8 times higher if participants' tested negative for THC-COOH than if they tested positive (OR = 8.2; 95% CI: 2.6, 27.0; $P < 0.001$).

© 2019 American Medical Association. All rights reserved.

eReferences

Baker NL, Gray KM, Sherman BJ, Morella K, Sahlem GL, Wagner AM, McRae-Clark AL. *Drug Alcohol Depend* 2018; 187: 270-277.

van Buuren S, Groothuis-Oudshoorn. Mice: Multivariate imputation by chained equations in R. *J. Stat. Softw.* 2011; 45(3):1-67.

Rubin D. *Mutliple Imputation for Nonresponse in Surveys.* New York: John Wiley & Sons; 1987

Schafer JL. *Analysis of incomplete multivariate data*. London: Chapman & Hall; 1997.

Vuong QH. Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica.* 1989; 57(2): 307-333.

© 2019 American Medical Association. All rights reserved.