Early Cannabis Use and Neurocognitive Risk: A Prospective Cohort Functional Magnetic Resonance Imaging Study

Supplemental Information

Cannabis Use Assessment

Cannabis use was assessed at 12-, 13-, 14-, and 15-years old. At each visit, cannabis age-of-onset ("how old were you when you tried marijuana?"), past-year frequency (every day, almost every day, 3-4 days per week, 1-2 days per week, 2-3 days per month, once per month, 6-11 days per year, 1-5 days per year, no use) and quantity ("on the days when you use(d) marijuana, about how many joints do (did) you usually smoke?") were assessed. As in previous work from the MHCPD (1,2), consumption from different methods (i.e., Blunt, Hashish) was converted to an equivalency of marijuana joints based on THC estimates from previous literature (3). Our cannabis dosage estimate combined frequency and quantity measures (the product of average frequency and quantity) to create an average daily dose of cannabis (joints-per-day) for each assessment period.

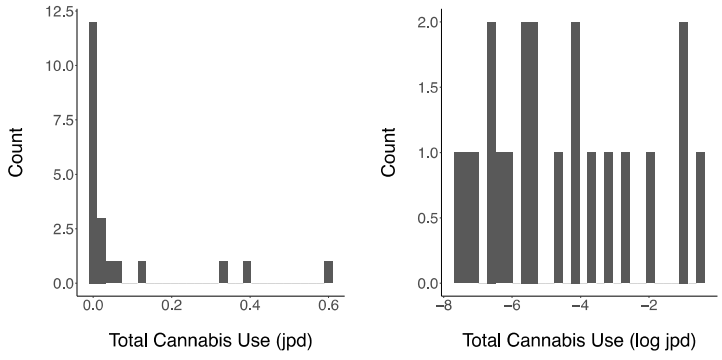

Figure S1. Raw (left) and log (transformed) total cannabis use distribution in user group with reported dose (n = 21). Total cannabis use calculated as the sum of joints-per-day (jpd) at 13-, 14- and 15-years old (range, .005 - .6 jpd).

fMRI Data Acquisition and Preprocessing

Data Acquisition. Imaging data were collected using a 3.0-T Siemens Magnetom TIM Trio at the Magnetic Resonance Research Center at the University of Pittsburgh. Structural images used for functional registration and conversion to a standardized template were collected using a magnetization prepared rapid acquisition gradient-echo (MP-rage) pulse sequence with 192 slices (1mm slice thickness; 1mm isotropic voxels). Functional data were collected using an echo-planar imaging (EPI) sequence with the following parameters: TR=2.0s, TE=20ms, Flip Angle=80˚, and 128 x 120 acquisition matrix with a field of view of 220mm. Thirty-three slices were collected in the axial plane with an anisotropic voxel size of 1.72mm x 1.72mm x 3mm and .75mm gap between slices covering the entire cortex and some of the cerebellum.

Preprocessing. Standardized preprocessing procedures were used and utilized the same pipeline as recent work from our group (4, 5), employing tools from Analysis and Visualization of Functional Neuroimages (AFNI, Bethesda, MD) and FSL (FMRIB, Oxford, UK). Preprocessing procedures included wavelet despiking (AFNI 3dDespike), slice timing correction, motion correction (mcflirt; (6)), brain extraction, non-linear registration of functional data to a standardized anatomical brain (3mm MNI-152 template: 2009c), spatial smoothing with FWHM of 5mm (SUSAN; (7)), high pass filtering at 80 volumes (.00625Hz), and scaling by 10,000 of the global median.

Additional Information on Measures from Cognitive Battery

WISC-IV. All factor scores were estimated according to WISC-IV scoring manual. Perceptual Reasoning (PER): this factor includes subtests designed to asses nonverbal problem solving and reasoning, including Block Design, Picture Concepts, and Matrix Reasoning. Processing Speed (PRO): this factor includes subtests designed to assess the speed of information processing, including Coding and Symbol Search. Verbal Comprehension (VER): this factor includes Similarities, Vocabulary, and Comprehension subtests, designed to measure general verbal skills. Working Memory (WM): this factor includes subtests designed to assess working memory, including Digit Span and Letter-Number Sequencing. Compared to the scanner working memory task and other tests in our cognitive battery, WISC-IV working memory tasks are assessed verbally and thus do not include a visuospatial component. The four factor structure of the WISC-IV has been shown to be invariant between clinical and nonclinical samples (8).

CANTAB. CANTAB, computer-based cognitive testing was performed in accordance with standard procedures from the distributor. Participants performed four tests: Intra-Extra Dimensional Set Shift (IED), One Touch Stockings of Cambridge (OTS), Spatial Span (SSP), and Spatial Working Memory (SWM). IED was designed to test attentional set shifting and flexibility. Participants must incorporate feedback to learn which of two targets are “correct” and what features (colored shapes, white lines) distinguished them. After six correct trials the rules are changed and new features are enforced. In this manner, the IED shares features to the Wisconsin Card Sort . OTS was

designed to test executive function and planning that is based upon the Tower of Hanoi/London Tests. Participants must move a set of disks to match a visual sample, in the fewest number of moves. SSP was designed to test visuospatial working memory. Participants must repeat a visual sequence (squares that briefly change color). SWM was also designed to test visuospatial working memory. Participants must use process of elimination to find "tokens" from a series of visual target locations. "Tokens" cannot occur in the same location more than once in a trial, such that participants must remember their previous selections. More information and example stimuli from the CANTAB tasks can be found at http://www.cambridgecognition.com/cantab/cognitive-tests/.

In the current analysis, primary performance measures were analyzed for each CANTAB test (IED: total errors adjusted; OTS: problems solved on first choice; SSP: span length; SWM: total errors). As group differences were observed on CANTAB OTS, additional analysis was performed on this measure (Figure S2, below). Final scores from IED and SWM were reverse scored so that higher scores represented better performance on all tests.

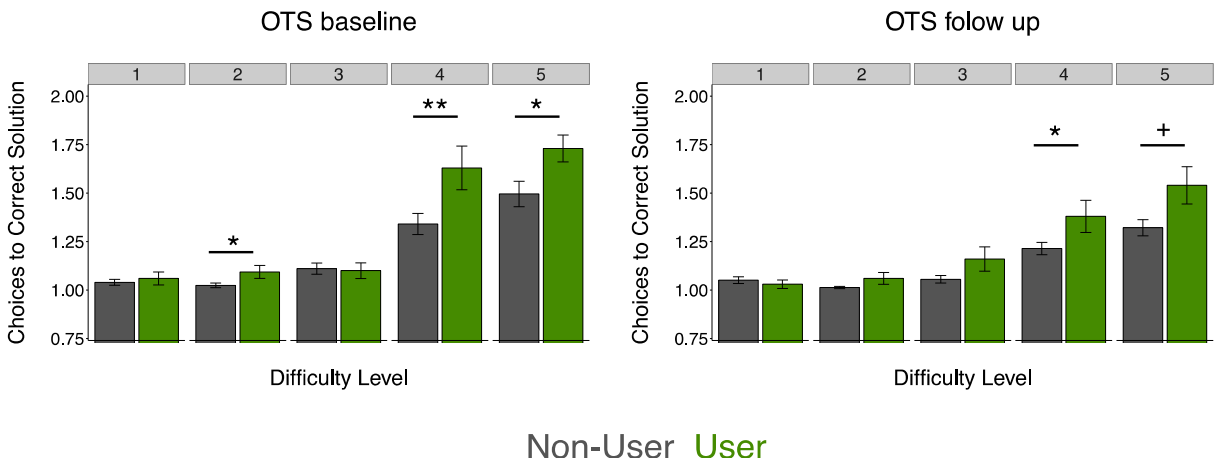

Figure S2. Group means for individual difficulty levels of One Touch Stockings of Cambridge (OTS; CANTAB Battery) at baseline (left) and follow up (right). *Note.* ** p < .01 , * p < .05, + p = .067.

Additional Scanner Accuracy Modeling Information

Mean performance measures across each combination of the 2 (load) x 2 (delay length) x 2 (cue validity) conditions (eight means) were calculated for each participant at each visit. Mixed-effects models (lme4 package; (9)) with maximum likelihood estimation were used to examine main effects of, and interactions between, task conditions, cannabis use, and visit as well as the role of covariates (SES composite and CBCL externalizing scores). Fixed effects were included in two phases: primary models, which only included the task conditions, visit (where relevant), and cannabis group, and covariate models, which additionally included the SES composite or CBCL externalizing scores. CBCL externalizing scores were treated as time-varying covariates, where 12- and 15-year old scores were included at the appropriate visit. Fixed effect significance values were obtained through the car package (chi-square test; (10)). Simple effects were obtained through the lsmeans package (11). Cook's distance and dfbetas (both

cutoffs >1) were used to examine potential influential observations. No participants passed these thresholds. Transformations of accuracy (arcsine) and reaction time (log) data were explored, however the pattern of significance was unchanged.

Scanner Working Memory Accuracy Task Effects and Interactions

Given the similarity of group differences in the full sample and longitudinal sample (see Results) and the added statistical power utilizing both baseline and follow up data, we report task effects and interactions within the longitudinal sample only.

Task Effects. Only the WM load condition was associated with WM accuracy differences (Table S1).

Table S1. WM Accuracy Means by Task Condition

	Mean Accuracy Condition A	**Mean Accuracy Condition B**	**Condition A vs. Condition B**
Load	*1 location*	*3 locations*	$\chi^2_{(1)} = 149.73$, t = -12.24, p < .001
	86.2 %	93.6 %	
Delay Length	1500ms delay	6000ms delay	$\chi^2_{(1)} = 0.41$, t = -0.64, p = .521
	89.7 %	90.1 %	
Cue Validity	Non-Match	Match	$\chi^2_{(1)} = 1.80$, t = -1.34, p = .179
	89.5 %	90.3 %	

Note. Mean accuracy estimates (% correct) are least-squares means from an additive model with usage group, visit, WM load, delay, and cue validity. Random intercepts were estimated for each subject. Significance testing utilized Wald's chi-square test.

Cannabis Group Interactions with WM Load. We restricted our analysis of cannabis group task interactions to WM load, as this was the only significant task factor for

accuracy. Averaging across baseline and follow visits, cannabis group accuracy differences did not vary by WM load (usage group by load interaction: $\chi^2_{(1)} = 0.64$, t = -0.80, p = .424). However, a three-way interaction between cannabis group, WM load, and visit was significant ($\chi^2_{(2)} = 16.25$, $p < .001$). Post-hoc testing revealed at baseline, those who would go onto use cannabis by 15 (follow up) had significantly lower accuracy in the low load/1 location condition ($t_{(144.06)} = -2.26$, p = .026) but not in the high load/3 locations condition ($t_{(144.06)} = -1.54$, p = .125). At follow up, group differences remained significant for the low load/1 location condition ($t_{(144.06)} = -2.20$, p = .030) but were at a trend for the high load/3 locations condition ($t_{(144.06)} = -1.88$, p = .063).

Scanner Working Memory Reaction Time Task Effects and Interactions

As in the accuracy analysis, we report task effects and interactions for reaction time within the longitudinal sample only.

Task Effects. The WM load, delay time, and cue validity conditions were each associated with WM reaction time differences (Table S2).

Table S2. WM Reaction Time Means by Task Condition

	Mean Accuracy Condition A	Mean Accuracy Condition B	Condition A vs. Condition B
Load	*1 location*	*3 locations*	$\chi^2_{(1)} = 575.49$, t = -23.99, p < .001
	985.58 ms	1160.23 ms	
Delay Length	1500ms delay	6000ms delay	$\chi^2_{(1)} = 9.72$, t = 3.12, p = .002
	1084.26 ms	1061.55 ms	
Cue Validity	Non-Match	Match	$\chi^2_{(1)} = 82.72$, t = 9.10, p < .001
	1106.01 ms	1039.80 ms	

Note. Mean reaction time estimates (A & B) are least-squares means from an additive model with usage group, visit, WM load, delay, and cue validity. Random intercepts were estimated for each subject. Significance testing utilized Wald's chi-square test.

Cannabis Group Interactions with WM Load. Averaging across baseline and follow visits, cannabis group reaction time differences did not vary by WM load (usage group by load interaction: $\chi^2_{(1)} = 0.01$, t = 0.09, p = .931), delay time (usage group by delay time interaction: $\chi^2_{(1)} = 1.74$, t = 1.32, p = .187), or cue validity ($\chi^2_{(1)} = 0.43$, t = -0.65, p = .514). A significant three-way interaction for cannabis group, visit, and WM load was significant ($\chi^2_{(2)} = 6.02$, p = .049). However, post-hoc testing revealed non-significant group differences for both loads at both visits (p's > .224). Three-way interactions with cannabis group, visit, and delay length ($\chi^2_{(2)} = 2.02$, p = .364) and cannabis group, visit, and cue validity were not significant ($\chi^2_{(2)} = 1.69$, p = .429).

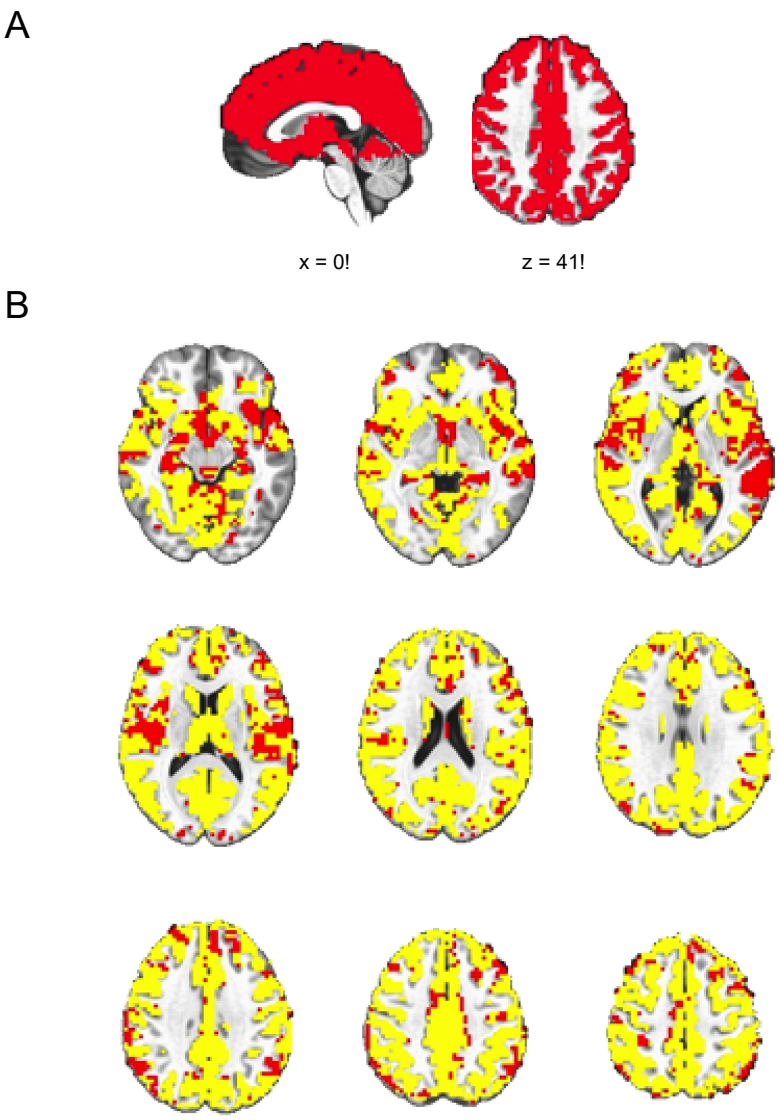

Figure S3. Voxel Inclusion. **A**) Voxels included in the full coverage and grey-matter mask are displayed in red. Voxels not included in this mask were excluded from all analyses. **B)** Voxels within the full coverage mask that have a main effect of TR ($p < .005$, uncorrected) are displayed in yellow. Displayed voxels are from longitudinal model (included voxels = 20,451) and were similar to baseline (included voxels = 19,180) and follow up (included voxels = 15,727) voxel inclusion masks.

Additional Information on Cluster Correction

Based on recommendations by AFNI software, cluster correction was performed using the acf option in AFNI's 3dClustSim program. In order to have the most parsimonious and precise smoothness estimates as inputs to 3dClustSim, the grand mean acf parameters (across subjects and time points) from GLM-1 residuals was used (parameters a (0.684), c (3.200), f (13.215) ; equivalent FWHMX (8.71 mm)). Additional specifications for cluster correction included: the requirement that voxel faces must touch (nearest neighbor 1 option) and 1-sided testing, as all voxelwise tests were F-statistics. Based on these parameters, 3dClustsim returns a table of corrected significance value for varying voxelwise p-value thresholds. We used voxelwise FDR correction to select this p-value, such that any result cluster had the added interpretation that all individual voxels had corrected significance. Accordingly, our correction approach utilized a two-step correction procedure.

Scanner WM Accuracy Covariate Analysis

In the longitudinal sample, accuracy differences between usage groups remained significant when covarying EXT (baseline: $t_{(93.58)}$=2.15, p=.034; follow-up: $t_{(99.59)}$=2.38, p =.019) and EXT was not a significant predictor of WM accuracy in this model (EXT: $\chi^2_{(1)}$=0.06, t=0.25, p=.800). While covarying SES, the difference between users and non-users in WM accuracy was reduced to a trend at baseline ($t_{(95.44)}$=1.82, p=.072) and follow-up ($t_{(95.44)}$=2.09, p=.051). However, SES was not a significant predictor of scanner WM accuracy ($\chi^2_{(1)}$=2.22, t=1.49, p=.136), ruling out potential mediation.

Accuracy difference between usage groups remained significant when covarying

alcohol usage group (baseline: $t_{(92.69)}$=2.72, p=.008; follow-up: $t_{(92.69)}$=2.38, p =.005). Within this model, alcohol usage group had a significant main effect (ALC: $\chi^2_{(1)}$=3.89, t=-1.97, p=.048). However, alcohol usage group was not significantly associated with WM accuracy when cannabis usage group was not covaried (ALC: $\chi^2_{(1)}$=0.59, t=-0.77, p=.441). Given the high overlap between the two groups and concerns with multicollinearity, we do not interpret this as a significant association with alcohol usage group. The interaction between alcohol usage group and cannabis usage group was not significant ($\chi^2_{(1)}$=0.43, t=-0.66, p=.510), indicating no additive effects of alcohol and cannabis usage groups.

Head Motion

Head motion did not significantly differ between groups at either baseline (user vs. non-user, mean Euclidean norm (enorm): $t_{(32.90)}$=-.030, p =.976, d=-.008; percentage of censored tr's (censor): $t_{(32.61)}$=.141, p =.889, d=.037) or follow-up (user vs. non-user, enorm: $t_{(32.90)}$=-1.31, p =.198, d=.326; censor: $t_{(58.47)}$=-1.18, p =.243, d=.226). Motion significantly decreased between visits (main effect of visit, enorm: $\chi^2_{(1)}$=39.26, t=-6.27, p < .001; censor: $\chi^2_{(1)}$=13.52, t=-3.68, p < .001) but the change between time points did not differ by group (group by visit interaction, enorm: $\chi^2_{(1)}$=1.58, t=1.26, p=.208; censor: $\chi^2_{(1)}$=1.51, t=1.23, p=.219). Further, within the cannabis group with a dosage estimate (baseline: n=20; follow-up: n=14), motion was not associated with cannabis dose at baseline (enorm: β=.079, t_r=.514, p=.622; censor: β=-.002, t_r=-.025, p=.981) or follow-up (enorm: β=-.055, t_r=-.171, p=.866; censor: β=.036, t_r=.611, p=.569).

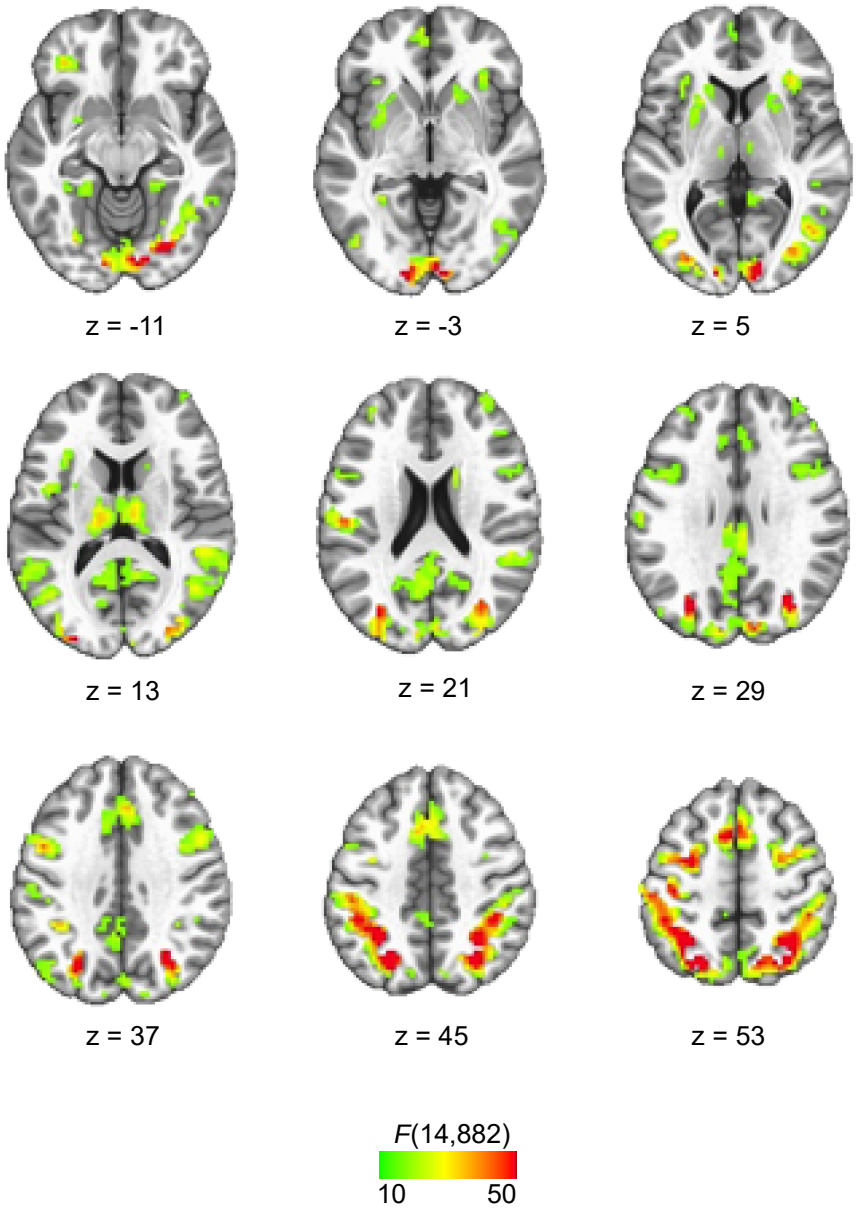

Figure S4. Main Effect of TR while Covarying Visit in Longitudinal Sample. Large F-values suggest voxel HRF time series significantly varied in response to the WM task. Statistical maps displayed over MNI-152 template in neurological view.

Baseline fMRI Group Differences in Subsamples

To examine the impact of attrition and inconsistent cannabis report at baseline, we repeated voxelwise testing at baseline with the longitudinal sample and the full sample when removing the four subjects with inconsistent onset. Analysis utilized the same methods as reported in the methods section in the main text. We then performed a conjunction analysis (Figure S5) of group differences in the three samples (full sample, longitudinal sample, full sample removing inconsistent onset). This revealed consistent overlap in the location of group differences in all three samples.

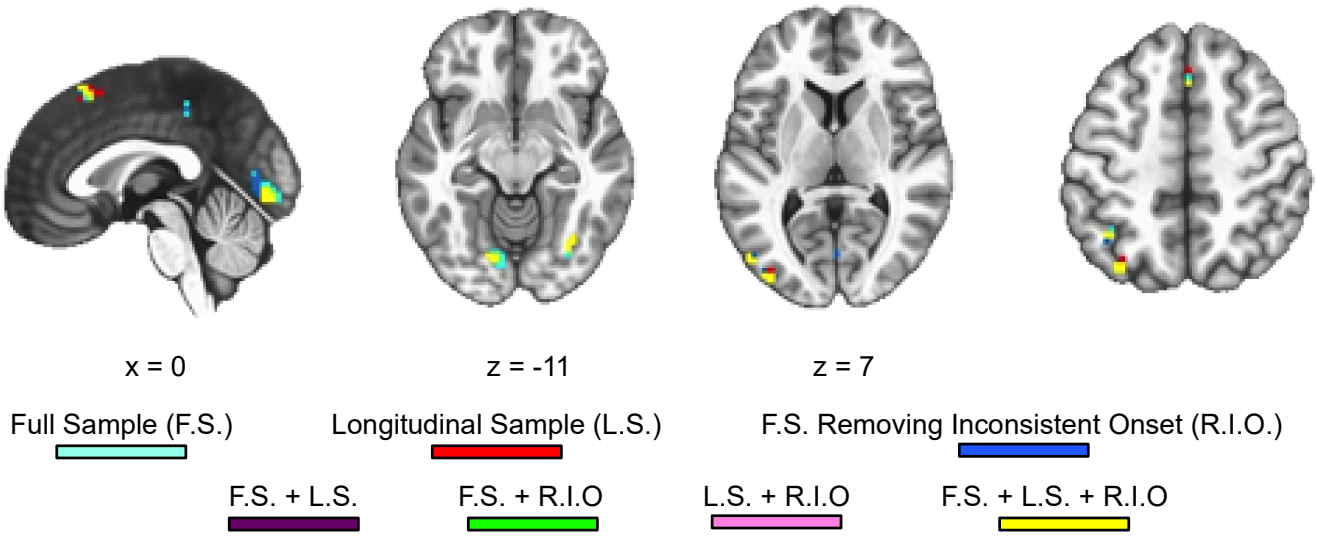

Figure S5. Conjunction analysis of group differences in Full Sample (F.S.), Longitudinal Sample (L.S.), and Full Sample Removing Subjects with Inconsistent Cannabis Onset Report (R.I.O.)($q < .05$, 11 or more contiguous voxels, faces touching). Statistical maps displayed over MNI-152 template in neurological view.

We next examined whether group differences in the clusters reported for the full sample were disproportionality affected by subjects with attrition or inconsistent cannabis onset report. In particular, we examined whether removing the specific

subjects from the subsamples influenced the observed effect sizes more than any random set of the same number of subjects. To test this, we compared the observed subsamples' effect sizes (Cohen's d from BOLD Dot Product) to distributions of effect sizes that were created from permutations of the original sample that removed the same number of subjects at random. This procedure revealed that effect sizes were significantly larger in the longitudinal sample (compared to the full sample) in the cingulate and lateral occipital gyrus. Removing the four subjects with inconsistent cannabis onset did not exceed what would be expected by chance for any four random users. See Figure S6 for effect sizes in subsamples and p-values from permutation tests.

Region	**d F.S.**	**d L.S.**	**d R.I.O.**	**p-value** d F.S. > d L.S.	**p-value** d F.S. > d R.I.O.
R- Lingual Gyrus (A)	.871	.945	.730	.319	.936
L- Lingual Gyrus	.778	.799	.698	.442	.775
L- Precuneus	-.723	-.644	-.626	.670	.857
L- MFG	-.905	-.975	-.883	.358	.604
L- IPL	-.598	-.656	-.653	.408	.376
R- MFG	-1.168	-1.046	-1.039	.767	.885
Cingulate	.016	-.545	.024	.001	.793
pre-SMA	-.899	-.927	-.866	.450	.639
R-Lingual Gyrus (B)	.883	.832	.826	.591	.599
L-Lateral Occipital Gyrus (A)	.847	.857	.754	.441	.816
L-Lateral Occipital Gyrus (B)	.841	1.132	.790	.019	.678

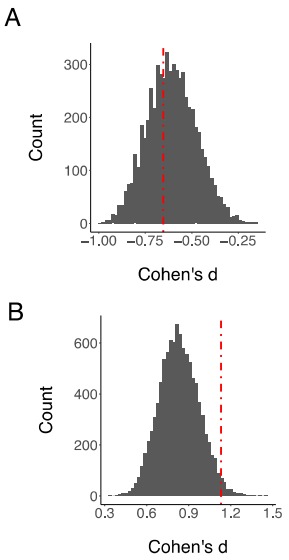

Figure S6. Left, Effect sizes (Cohen's d from BOLD Dot Product) in the full sample (F.S.), longitudinal sample (L.S.) and full sample removing subjects with inconsistent onset (R.I.O.). P-value d F.S. > L.S., two-tailed significance value from permutation test comparing observed L.S. effect size to 10,000 random subsamples with the same number of subjects (7 users, 12 non-users) excluded as the L.S. P-value d F.S. > R.I.O., two-tailed significance value from permutation test comparing observed R.I.O effect size to all permutations of removing four user subjects (5,985). **Right**, Examples of non-significant (A; L-IPL) and significant permutation tests (B; Lateral Occipital Gyrus (B)). Observed effect size displayed as red hashed line. Null distribution displayed as grey histogram.

Main Effects of Usage Group in fMRI Analysis

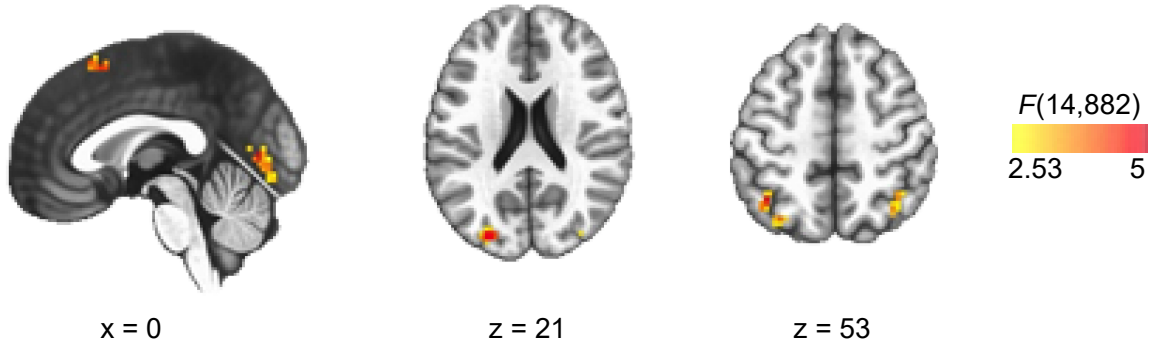

Figure S7. Clusters with Significant Group by TR interactions while Covarying Visit in Longitudinal Sample (q < .05, 11 or more contiguous (faces touching) voxels). Statistical maps displayed over MNI-152 template in neurological view.

Table S3. Simple Effects of Usage Group in fMRI Analysis from Longitudinal Sample

Region	BA	F-Value (Peak)	X	Y	Z	Total Voxels	q < .05 Baseline Voxels	q < .05 Follow Up Voxels	d Baseline Non User > User	d Follow up Non User > User
R-Lingual Gryus (A)	18	5.89	-1.5	73.5	7.5	43	**29***	**23***	.206	-.144
L-Lingual Gyrus	18	3.73	16.5	67.5	-7.5	35	**21***	5	-.092	-.299
L- Lateral Occipital Gyrus	19	6.26	43.5	88.5	7.5	29	**25***	**25***	1.054	.739
L- Superior Parietal Lobule	7	5.26	25.5	67.5	46.5	29	**22***	**16***	-.797	-.616
L- Cuneus	19	6.46	25.5	88.5	22.5	24	**19***	**20***	.716	.904
L- Inferior Parietal Lobule	40	4.97	40.5	58.5	55.5	19	**15***	**13***	-.736	-.811
R- Cuneus	19	5.85	-19.5	88.5	37.5	18	2	18	-.516	-1.282
R- Inferior/ Superior Parietal Lobule	19/ 40	4.40	-37.5	55.5	55.5	17	**15***	4	-.879	-.574
R-Lingual Gyrus (B)	18	6.18	-28.5	73.5	-13.5	16	9	**15***	.714	.922
R-Middle Occipital Gyrus	19	4.49	-34.5	82.5	16.5	13	5	**13***	.601	1.029
Pre-SMA	8/6	4.45	-1.5	-16.5	55.5	12	**10***	6	-.866	-.732

Note. BA, Brodmann areas; F-value, cluster peak test statistic for group by TR interaction

across both baseline and follow up: Main Effect of Usage Group, *F*(14,882). X,Y,Z MNI-152 coordinates at peak. Total Voxels, total voxels in the cluster. q < .05 Baseline Voxels, number of significant (fdr-corrected across all voxels with Main effect of Usage Group) voxels at baseline. q < .05 Follow Up Voxels, number of significant (fdr-corrected) voxels at follow up. Clusters marked with * and bolded have significant simple effects at baseline or follow up based on cluster correction within the Main effect of Usage Group Mask (6 or more contiguous (faces touching) voxels, post-hoc q's < .05 (baseline: p = .039; follow up p = .034). d, Cohen's D from BOLD dot product at baseline and follow up from whole cluster.

Table S4. Socioeconomic Status Indirect Effects on BOLD Activation at Baseline

Region	**SES Indirect Pathway** β	**95 % C.I. Low** β	**95 % C.I. High** β
Baseline			
R- Lingual Gyrus (A)	-.057	-.179	.002
L- Lingual Gyrus	-.015	-.098	.046
L- Precuneus	-.036	-.147	.040
L- Middle Frontal Gyrus (MFG)	.034	-.021	.109
L- Inferior Parietal Lobule (IPL)	.016	-.068	.091
R- Middle Frontal Gyrus (MFG)	.027	-.045	.098
Paracentral Lobule/Cingulate Gyrus	.034	-.022	.147
pre-Supplementary Motor Area (SMA)	.059	-.020	.162
R-Lingual Gyrus (B)	-.043	-.161	.034
L-Lateral Occipital Gyrus (A)	-.032	-.168	.025
L-Lateral Occipital Gyrus (B)	-.053	-.151	.033
Follow-Up			
R-Cuneus (A)	-.006	-.073	.139
R-Cuneus (B)	-.061	-.187	.043
R-Cuneus (C)	-.002	-.063	.091
R-Cuneus Aggregate (A,B,C)^{L}	-.033	-.122	.092
Visit by Group			
Posterior Cingulate^{L}			
Baseline	.068	-.011	.178
Follow-up	-.004	-.077	.067

Note. β, standardized indirect effect (a*b pathway) from cannabis usage group through SES to BOLD activation differences at baseline. 95% C.I., 95% confidence interval based on 5,000 draws in bootstrap procedure (mediation package; (12)). ^{L}, estimates based on the longitudinal sample only.

Table S5. Alcohol Associations with BOLD Activation in Clusters Defined by Cannabis Group

Region	Non-Alcohol User vs. Alcohol User	Alcohol User by Cannabis User Interaction	Total Alcohol Use
	d	*F*	β
Baseline			
R- Lingual Gyrus (A)	-.407	1.41	-.043
L- Lingual Gyrus	-.322	1.56	.103
L- Precuneus	-.046	1.17	.115
L- MFG	.035	0.51	.565
L-IPL	.004	0.00	.177
R- MFG	.435	0.38	.303
Cingulate	.149	0.67	.083
pre-SMA	.479	1.30	.050
R-Lingual Gyrus (B)	.130	0.01	.006
L-Lateral Occipital Gyrus/BA 19 (A)	-.162	0.10	.129
L-Lateral Occipital Gyrus/BA 18 (B)	-.089	0.11	-.011
Follow-up			
R-Cuneus (A)	-.427	1.09	-.688
R-Cuneus (B)	-.360	1.77	-.228
R-Cuneus (C)	-.027	0.05	.050
R-Cuneus Aggregate (A,B,C)[L]	-.330	0.02	-.024
Visit by Group			
Posterior Cingulate[L]			
Baseline	.437	0.00	-.316
Follow-Up	.083	0.71	-.075

We present three analyses to demonstrate clusters defined by cannabis group were not biased by the alcohol usage. First, to determine whether there was evidence of an independent effect of alcohol use in the cluster, we examined group differences in alcohol usage group, while covarying cannabis group (the feature used to define the cluster). Next, we examined the interaction between alcohol and cannabis usage groups to determine whether participants with combined cannabis and alcohol usage did not bias these clusters. Finally, to mirror our analysis of dimensional cannabis dose, we examined the continuous association between BOLD activation and total alcohol use in all reported clusters. No significant (corrected) results were found.

Note. d, Cohen's d from BOLD dot product, covarying cannabis usage group. F, Interaction statistic between alcohol usage group and cannabis usage group. β, standardized regression coefficient predicting BOLD dot product from total alcohol use (model run only in those with reported alcohol use). [L], estimates based on the longitudinal sample only.

Voxelwise Covariate Conjunction

In order to assess overlap between voxelwise cannabis usage group results and voxelwise SES, EXT, and alcohol usage group results, we performed the same analysis for each of these covariates (3dMVM, q < .05, 11 or more contiguous voxels) at baseline and follow up visits, and performed a conjunction analysis. Generally, very minimal overlap was observed in corrected, significant voxels. At baseline 18 voxels had overlapping significance for cannabis usage group and SES in the lingual gyrus (Figure S8, Left). At follow up, 5 voxels had overlapping significance for cannabis usage group and EXT in the cuneus (Figure S8, Right). No overlap in significant voxels was observed for cannabis usage group and these variables at the other time point (i.e., SES at follow-up; EXT at baseline). No overlap in significant voxels was observed for cannabis usage group and alcohol usage group at either time point.

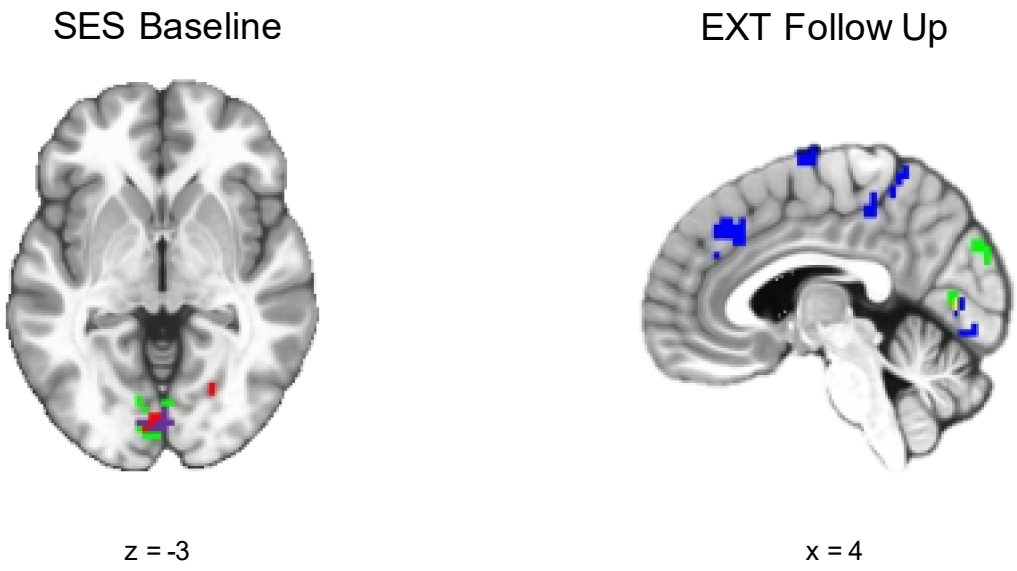

Figure S8. Left, significant voxels from cannabis usage group (green) and SES (red) and their overlap (purple) at the baseline visit. **Right,** significant voxels from cannabis usage group (green) and EXT (blue) and their overlap (yellow). Statistical maps displayed over MNI-152 template in neurological view.

Left Lateral Occipital BOLD Mediates Group Working Memory Accuracy Differences at Baseline

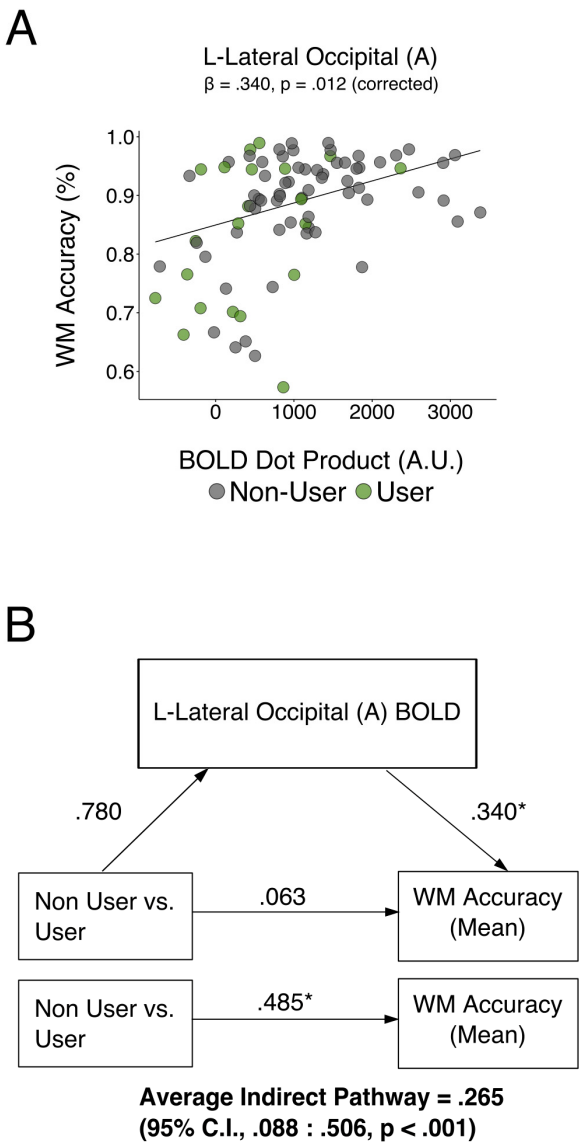

Figure S9. A) Significant association between L-lateral occipital activation and working memory accuracy, while covarying usage group. **B**) Brain-behavior mediation model.

Table S6. Cuneus Aggregate (A,B,C) BOLD Associations with Behavioral Battery

	β	t_r	p
Baseline			
FSIQ	.145	1.13	.500
PER	.278	2.33	.119
PRO	.099	0.77	.593
VER	-.001	-0.01	.995
WM	.019	0.14	.995
IED	.137	1.10	.500
OTS	.221	1.88	.193
SSP	-.100	-0.74	.593
SWM	.307	2.48	.119
Follow Up			
FSIQ	-.219	-1.31	.436
PER	-.069	-0.42	.766
PRO	-.220	-1.41	.436
VER	-.070	-0.42	.766
WM	-.379	-2.34	.185
IED	.000	0.00	.998
OTS	.241	1.57	.436
SSP	.124	0.72	.766
SWM	-.064	-0.46	.766
Difference Scores (Follow Up – Baseline)			
FSIQ	-.163	-1.09	.620
PER	-.103	-0.70	.691
PRO	-.156	-1.12	.620
VER	-.026	-0.17	.864
WM	-.090	-0.62	.691
IED	.045	0.39	.770
OTS	.241	1.92	.391
SSP	-.108	-0.67	.691
SWM	.230	1.77	.391

Note. β, standardized regression coefficient from m-estimation linear model; t_r = robust t-statistic from m-estimation linear model; p, fdr-corrected p-value. WISC-IV: FSIQ, Full-scale IQ; PER, Perceptual Reasoning; PRO, Processing Speed; VER, Verbal Reasoning; WM, Working Memory. CANTAB: IED, Intra-Extra Dimensional Set Shift, OTS, One Touch Stockings of Cambridge, SSP, Spatial Span; SWM, Spatial Working Memory. Difference scores for dot-products were first z-scored at each visit.

Table S7. Cuneus (A) BOLD Associations with Behavioral Battery

	β	t_r	p
Baseline			
FSIQ	.053	0.41	.779
PER	.166	1.37	.695
PRO	.113	0.87	.695
VER	-.051	-0.40	.779
WM	-.085	-0.68	.749
IED	.110	0.88	.695
OTS	.029	0.24	.814
SSP	-.143	-1.06	.695
SWM	.172	1.29	.695
Follow Up			
FSIQ	-.201	-1.27	.472
PER	.023	0.13	.980
PRO	-.243	-1.61	.332
VER	-.059	-0.37	.980
WM	-.360	-2.43	.145
IED	.008	0.05	.980
OTS	.279	1.93	.293
SSP	-.004	-0.03	.980
SWM	.051	0.38	.980
Difference Scores (Follow Up – Baseline)			
FSIQ	-.053	-0.36	.893
PER	-.019	-0.13	.893
PRO	-.086	-0.64	.893
VER	-.136	-1.00	.893
WM	.031	0.22	.893
IED	.171	1.50	.801
OTS	.089	0.71	.893
SSP	-.182	-1.34	.801
SWM	.033	0.27	.893

Note. β, standardized regression coefficient from m-estimation linear model; t_r = robust t-statistic from m-estimation linear model; p, fdr-corrected p-value. WISC-IV: FSIQ, Full-scale IQ; PER, Perceptual Reasoning; PRO, Processing Speed; VER, Verbal Reasoning; WM, Working Memory. CANTAB: IED, Intra-Extra Dimensional Set Shift, OTS, One Touch Stockings of Cambridge, SSP, Spatial Span; SWM, Spatial Working Memory. Difference scores for dot-products were first z-scored at each visit.

Table S8. Cuneus (B) BOLD Associations with Behavioral Battery

	β	t_r	p
Baseline			
FSIQ	.078	0.58	.748
PER	.092	0.72	.748
PRO	.074	0.55	.748
VER	-.051	-0.38	.790
WM	.148	1.17	.719
IED	-.012	-0.09	.929
OTS	.253	2.15	.314
SSP	.135	1.00	.719
SWM	.194	1.42	.719
Follow Up			
FSIQ	-.059	-0.40	.963
PER	-.043	-0.28	.963
PRO	.007	0.05	.963
VER	-.062	-0.42	.963
WM	-.162	-1.09	.963
IED	-.042	-0.27	.963
OTS	.083	0.60	.963
SSP	.184	1.20	.963
SWM	.006	0.05	.963
Difference Scores (Follow Up – Baseline)			
FSIQ	-.188	-1.34	.588
PER	-.062	-0.42	.868
PRO	-.067	-0.48	.868
VER	-.002	-0.01	.991
WM	-.132	-0.92	.810
IED	.059	0.48	.868
OTS	.165	1.29	.588
SSP	.014	0.09	.991
SWM	.188	1.46	.588

Note. β, standardized regression coefficient from m-estimation linear model; t_r = robust t-statistic from m-estimation linear model; p, fdr-corrected p-value. WISC-IV: FSIQ, Full-scale IQ; PER, Perceptual Reasoning; PRO, Processing Speed; VER, Verbal Reasoning; WM, Working Memory. CANTAB: IED, Intra-Extra Dimensional Set Shift, OTS, One Touch Stockings of Cambridge, SSP, Spatial Span; SWM, Spatial Working Memory. Difference scores for dot-products were first z-scored at each visit.

Table S9. Cuneus (C) BOLD Associations with Behavioral Battery

	β	t_r	p
Baseline			
FSIQ	-.027	-0.21	.870
PER	.145	1.16	.582
PRO	-.103	-0.77	.659
VER	-.021	-0.17	.870
WM	-.123	-0.98	.600
IED	.060	0.48	.813
OTS	.174	1.45	.452
SSP	-.238	-1.83	.362
SWM	.235	1.82	.362
Follow Up			
FSIQ	-.204	-1.33	.558
PER	-.141	-0.98	.608
PRO	-.226	-1.54	.558
VER	-.062	-0.41	.774
WM	-.316	-2.08	.366
IED	.007	0.05	.964
OTS	.161	1.12	.592
SSP	.108	0.72	.608
SWM	-.104	-0.82	.608
Difference Scores (Follow Up – Baseline)			
FSIQ	-.087	-0.60	.708
PER	-.202	-1.51	.403
PRO	-.102	-0.76	.708
VER	.116	0.79	.708
WM	-.091	-0.65	.708
IED	.010	0.09	.925
OTS	.242	2.01	.212
SSP	-.073	-0.47	.716
SWM	.280	2.32	.212

Note. β, standardized regression coefficient from m-estimation linear model; t_r = robust t-statistic from m-estimation linear model; p, fdr-corrected p-value. WISC-IV: FSIQ, Full-scale IQ; PER, Perceptual Reasoning; PRO, Processing Speed; VER, Verbal Reasoning; WM, Working Memory. CANTAB: IED, Intra-Extra Dimensional Set Shift, OTS, One Touch Stockings of Cambridge, SSP, Spatial Span; SWM, Spatial Working Memory. Difference scores for dot-products were first z-scored at each visit.

Table S10. Behavior Battery Results: Group Differences and Relationship with Socioeconomic Status (SES) Composite and Parent-reported Externalizing Symptoms (EXT).

	Non-User > User				SES	
	d	t_w	p	p SES cov	β	p
Baseline						
FSIQ	.456	1.678	.186	.194	.400	.001
PER	.741	2.913	.028	.102	.407	.001
PRO	.019	0.076	.940	.925	.202	.085
VER	.586	2.148	.099	.109	.324	.008
WM	-.185	-0.701	.549	.648	.150	.197
IED	.531	2.089	.099	.178	.200	.085
OTS	.868	3.080	.028	.034	.285	.010
SSP	.365	1.450	.200	.527	.328	.008
SWM	.366	1.491	.200	.444	.321	.009
Follow Up						
FSIQ	.512	1.985	.083	.175	.327	.018
PER	.702	3.241	.019	.078	.216	.074
PRO	.154	0.561	.744	.940	.266	.052
VER	.611	2.380	.061	.098	.355	.006
WM	-.025	-0.094	.979	.940	.040	.749
IED	.006	0.027	.979	.940	.134	.339
OTS	.831	2.623	.061	.078	.120	.325
SSP	.496	2.166	.065	.175	.238	.078
SWM	.594	2.308	.061	.119	.246	.052
Difference Scores (Follow up –Baseline)						
FSIQ	.249	0.883	.754	.807	-.048	.686
PER	.029	0.116	.908	.816	-.210	.267
PRO	.233	0.812	.754	.816	.215	.267
VER	.106	0.445	.754	.816	.106	.427
WM	.177	0.648	.754	.807	-.141	.427
IED	-.414	-1.570	.754	.807	-.045	.686
OTS	-.245	-0.838	.754	.816	-.172	.267
SSP	.110	0.430	.754	.807	-.143	.423
SWM	.199	0.741	.754	.816	-.103	.427

Note. **Left)** Usage group differences: d, Cohen's d; t_w = Welch's t-test; p, fdr-corrected p-value; p SES cov, fdr-corrected p-value for group difference while covarying SES; **Right)** SES β, standardized regression (m-estimation) coefficient predicting test performance from SES, while covarying usage group; p, fdr-corrected p-value from SES predicting test performance WISC-IV: FSIQ, Full-scale IQ; PER, Perceptual Reasoning; PRO, Processing Speed; VER, Verbal Reasoning; WM, Working Memory. CANTAB: IED, Intra-Extra Dimensional Set Shift, OTS, One Touch Stockings of Cambridge, SSP, Spatial Span; SWM, Spatial Working Memory.

Table S11. Behavior Battery Results: Relationships with Total Cannabis Use

	β	t_r	p
Baseline			
FSIQ	-.125	-0.546	.775
PER	-.209	-0.956	.775
PRO	-.220	-0.984	.775
VER	-.041	-0.172	.865
WM	.101	0.410	.775
IED	-.278	-1.262	.775
OTS	-.108	-0.456	.775
SSP	.169	0.757	.775
SWM	-.299	-1.358	.775
Follow Up			
FSIQ	-.086	-0.354	.972
PER	-.131	-0.536	.972
PRO	-.065	-0.257	.972
VER	-.153	-0.607	.972
WM	.154	0.569	.972
IED	.028	0.112	.972
OTS	.009	0.035	.972
SSP	.306	1.381	.972
SWM	-.181	-0.873	.972
Difference Scores (Follow Up – Baseline)			
FSIQ	.098	0.374	.836
PER	.049	0.208	.836
PRO	.211	0.910	.836
VER	-.211	-0.885	.836
WM	-.057	-0.229	.836
IED	.233	1.043	.836
OTS	.091	0.506	.836
SSP	.127	0.595	.836
SWM	.054	0.282	.836

Note. β, standardized regression coefficient from m-estimation linear model; t_r = robust t-statistic from m-estimation linear model; p, fdr-corrected p-value. WISC-IV: FSIQ, Full-scale IQ; PER, Perceptual Reasoning; PRO, Processing Speed; VER, Verbal Reasoning; WM, Working Memory. CANTAB: IED, Intra-Extra Dimensional Set Shift, OTS, One Touch Stockings of Cambridge, SSP, Spatial Span; SWM, Spatial Working Memory.

Supplemental References

1. Day N, Sambamoorthi U, Taylor P, Richardson G, Robles N, Jhon Y, *et al.* (1991): Prenatal marijuana use and neonatal outcome. *Neurotoxicol Teratol*. 13: 329–334.

2. Sonon KE, Richardson GA, Cornelius JR, Kim KH, Day NL (2015): Prenatal marijuana exposure predicts marijuana use in young adulthood. *Neurotoxicol Teratol*. 47: 10–15.

3. Gold MS (1989): *Marijuana*. New York, NY, US: Plenum Publishing Corporation.

4. Tervo-Clemmens B, Simmonds D, Calabro FJ, Day NL, Richardson GA, Luna B (2017): Adolescent cannabis use and brain systems supporting adult working memory encoding, maintenance, and retrieval. *NeuroImage*. .

5. Tervo-Clemmens B, Quach A, Luna B, Foran W, Chung T, De Bellis MD, Clark DB (2017): Neural Correlates of Rewarded Response Inhibition in Youth at Risk for Problematic Alcohol Use. *Front Behav Neurosci*. 11.

6. Jenkinson M, Bannister P, Brady M, Smith S (2002): Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. *NeuroImage*. 17: 825–841.

7. Smith S, Brady J (1997): SUSAN - a new approach to low level image processing. *Int J Comput Vis*. 23: 45–78.

8. Weiss LG, Keith TZ, Zhu J, Chen H (2013): WISC-IV and clinical validation of the four-and five-factor interpretative approaches. *J Psychoeduc Assess*. 31: 114–131.

9. Bates D, Maechler M, Bolker B (2013): *lme4: Linear mixed-effects models using S4 classes*. . Retrieved from http://CRAN.R-project.org/package=lme4.

10. Fox J, Weisberg S, Adler D, Bates D, Baud-Bovy G, Ellison S, *et al.* (2016): Package 'car.' *Companion Appl Regres R Package Version*. 2–1.

11. Lenth RV, Hervé M (2015): *lsmeans: Least-Squares Means, version 2.14*. .

12. Tingley D, Yamamoto T, Hirose K, Keele L, Imai K (2014): Mediation: R package for causal mediation analysis. Retrieved March 29, 2016, from http://dspace.mit.edu/handle/1721.1/91154.