ARTICLES
https://doi.org/10.1038/s41591-020-0806-7

nature medicine

Check for updates

A randomized proof-of-mechanism trial applying the 'fast-fail' approach to evaluating κ-opioid antagonism as a treatment for anhedonia

Andrew D. Krystal[1,2]✉, Diego A. Pizzagalli[3], Moria Smoski[2], Sanjay J. Mathew[4,5], John Nurnberger Jr[6], Sarah H. Lisanby[7], Dan Iosifescu[8], James W. Murrough[9], Hongqiu Yang[10], Richard D. Weiner[2], Joseph R. Calabrese[11], Gerard Sanacora[12], Gretchen Hermes[12], Richard S. E. Keefe[2], Allen Song[2], Wayne Goodman[4], Steven T. Szabo[2,13], Alexis E. Whitton[3,14], Keming Gao[11] and William Z. Potter[7]

The National Institute of Mental Health (NIMH) 'fast-fail' approach seeks to improve too-often-misleading early-phase drug development methods by incorporating biomarker-based proof-of-mechanism (POM) testing in phase 2a. This first comprehensive application of the fast-fail approach evaluated the potential of κ-opioid receptor (KOR) antagonism for treating anhedonia with a POM study determining whether robust target engagement favorably impacts the brain circuitry hypothesized to mediate clinical effects. Here we report the results from a multicenter, 8-week, double-blind, placebo-controlled, randomized trial in patients with anhedonia and a mood or anxiety disorder (selective KOR antagonist (JNJ-67953964, 10 mg; $n = 45$) and placebo ($n = 44$)). JNJ-67953964 significantly increased functional magnetic resonance imaging (fMRI) ventral striatum activation during reward anticipation (primary outcome) as compared to placebo (baseline-adjusted mean: JNJ-67953964, 0.72 (s.d. = 0.67); placebo, 0.33 (s.d. = 0.68); $F(1,86) = 5.58$, $P < 0.01$; effect size = 0.58 (95% confidence interval, 0.13–0.99)). JNJ-67953964, generally well tolerated, was not associated with any serious adverse events. This study supports proceeding with assessment of the clinical impact of target engagement and serves as a model for implementing the 'fast-fail' approach.

Clinical trials have documented the limited effectiveness of currently available medications for mood and anxiety spectrum disorders and speak to the need for new treatments[1–4]. Advances in research have provided unprecedented opportunities to meet this pressing need. Yet, the growing costs and failure rate associated with central nervous system (CNS) drug development have led many pharmaceutical companies to cease investing in CNS drugs, especially psychiatric medications[5–7].

In response to this alarming trend, the NIMH developed the New Experimental Medicine Studies: Fast-Fail Trials Program (https://www.nimh.nih.gov/research-priorities/research-initiatives/fast-fast-fail-trials.shtml) to target one factor identified as a key contributor to drug development failures: early-phase drug development methods that are notoriously unreliable and mislead companies into pursuing extremely costly, unsuccessful phase 3 studies[5–7]. The central contribution of the fast-fail approach is to include in all drug development efforts a rigorous determination of whether engaging the target of interest has the intended neurobiological effects. The assumption is that, by testing a very specific POM hypothesis, the development process has less vulnerability to bias and nonspecific effects[7,8]. Establishing POM decreases the likelihood that positive effects found in subsequent trials with clinical outcomes are due largely to factors other than the direct neural effects mediated by target engagement[8]. It also increases the confidence that negative results in phase 3 trials are interpretable, as indicating that achieving the hypothesized effect on brain function does not lead to the expected clinical effect. As a result, establishing POM is set as a precondition for proceeding to carry out such trials, which would otherwise be prohibitively risky and potentially uninterpretable.

Carrying out such POM studies requires a drug that is specific for the target, is safe and produces high target engagement, ideally established for a specific dose range with ligand-based positron emission tomography (PET)[8]. Most critically, POM studies require the use of biomarkers, which are assumed to be closer to pathophysiology and drug therapeutic mechanisms than clinical endpoints[5,9,10]. This allows for smaller phase 2 studies that more reliably predict phase 3 outcomes[5,9,10]. Additionally, under the fast-fail approach, use of the NIMH Research Domain Criteria Project (RDoC) framework is encouraged. Unlike the Diagnostic and Statistical Manual (DSM) diagnostic system, RDoC is based on neuroscience, thereby increasing the likelihood that the neural circuitry relevant to the condition of interest has been established and biomarkers exist, allowing confirmation that target engagement has the anticipated neural impact[8,11].

[1]Department of Psychiatry, University of California, San Francisco, San Francisco, CA, USA. [2]Departments of Psychiatry and Behavioral Sciences and Radiology, Duke University School of Medicine, Durham, NC, USA. [3]McLean Hospital, Harvard Medical School, Belmont, MA, USA. [4]Department of Psychiatry and Behavioral Sciences, Baylor College of Medicine, Houston, TX, USA. [5]Michael E. Debakey VA Medical Center, Houston, TX, USA. [6]Departments of Psychiatry and Neurobiology, Indiana University School of Medicine, Indianapolis, IN, USA. [7]National Institute of Mental Health, Bethesda, MD, USA. [8]Department of Psychiatry, New York University School of Medicine, New York, NY, USA. [9]Department of Psychiatry and Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [10]Duke Clinical Research Institute, Durham, NC, USA. [11]Department of Psychiatry, Case Western Reserve School of Medicine, Cleveland, OH, USA. [12]Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA. [13]Durham VA Health Care System, Durham, NC, USA. [14]The University of Sydney, Sydney, New South Wales, Australia. ✉e-mail: andrew.krystal@ucsf.edu

The first comprehensive application of this approach is presented here. Under the auspices of the NIMH New Experimental Medicine Studies: Fast-Fail Trials in Mood and Anxiety Spectrum Disorders (FAST-MAS) Program, we assessed the potential of KOR antagonism to have brain effects consistent with therapeutic benefit for anhedonia, a symptom of major depressive disorder (MDD) that cuts across traditional DSM as well as medical diagnoses[12]. Anhedonia can be defined in various ways and is not considered a diagnosis. Anhedonia, as broadly construed, is represented in several reward-related RDoC constructs ('reward responsiveness', 'reward learning' and 'reward valuation')[11,13]. Preclinical data strongly suggest that KOR antagonism will affect reward-related brain circuitry (in particular, ventral striatum) in a manner that could improve reward-associated function and reverse anhedonic symptoms and behaviors[13–21]. We selected and evaluated JNJ-67953964 (previously CERC-501 and LY2456302), a high-affinity, selective KOR antagonist with favorable pharmacologic and safety profiles, on the basis of (1) completed preclinical toxicology and human single- and multiple-ascending-dose studies[22,23] and (2) PET evidence that it robustly engages KOR at tolerated doses[24].

We based our decision to use reward-related activation in the ventral striatum as our primary target on the compelling preclinical literature indicating that KOR antagonism would release inhibition on dopamine (DA) neurons, increase nucleus accumbens function and prevent the development of anhedonic-like states[13,25]. Evidence suggests that KOR stimulation inhibits dopamine release in the striatum (nucleus accumbens) and induces a negative mood state[15]. Further, KOR agonists decrease phasic dopamine release in the nucleus accumbens and increase intracranial self-stimulation thresholds (an anhedonia model), whereas KOR antagonists have the opposite effect[14,17–19]. In light of evidence of positive correlations between reward-related dopamine release (as assessed by PET) and blood-oxygen-level-dependent (BOLD) activation (as assessed by fMRI)[26,27], reward-related BOLD activation in the ventral striatum was selected as the primary outcome variable (Fig. 1).

Results

Participants. A total of 163 participants were screened, and 94 met eligibility criteria (Extended Data Fig. 1). Of these, three withdrew consent before the baseline visit and two were unable to complete baseline procedures. Thus, 89 individuals were randomized, 45 to JNJ-67953964 and 44 to placebo. These 89 participants constituted the intention-to-treat (ITT) population. There were no significant differences ($P<0.05$) between the groups on any demographic or baseline variables in the ITT population (Table 1). For baseline characteristics of the ITT, as-treated, per-protocol and completer populations, see Supplementary Tables 1 and 2.

As-treated and per-protocol populations. Two participants were randomized to the JNJ-67953964 group but did not receive treatment and one participant in the placebo group received the wrong treatment, resulting in 43 participants receiving the assigned treatment in both groups. These 86 individuals constituted the as-treated population and the per-protocol population, as there were no major deviations from protocol.

Dropouts and the completer population. The completer population consisted of 33 participants in the JNJ-67953964 group and 35 participants in the placebo group. Ten participants in the JNJ-67953964 group dropped out. Two exited the study after double-blind treatment but before post-treatment follow-up. Eight participants in the placebo group dropped out during double-blind treatment and one exited between the end of double-blind treatment and post-treatment follow-up (for the reasons for dropout, see the Supplementary Note).

POM and efficacy. *Planned statistical analyses.* We carried out efficacy analyses in the ITT population and performed one-sided statistical tests using a *P*-value threshold of 5% for statistical significance. Analyses consisted of mixed-effects models including mean-centered baseline values, age, study site and sex as covariates. We computed JNJ-67953964 versus placebo group effect sizes by using Hedges' *g* (ref. [28]).

Primary outcome. Relative to placebo, JNJ-67953964 led to statistically significantly greater mean fMRI ventral striatal activation during anticipation of gain versus no-incentive trials in the Monetary Incentive Delay (MID) task ($F(1,86)=5.58$, $P<0.01$; Hedges' $g=0.58$, 95% confidence interval for effect size = [0.13, 0.99]; Table 2 and Fig. 2). Findings providing confirmatory support for this significant effect were that the between-group effect was statistically significant (1) with imputed results based on 20 imputations for missing data ($P<0.02$); (2) in the per-protocol and as-treated populations ($P<0.04$); (3) in the completer population ($P<0.05$); and (4) in analysis of the maximum fMRI ventral striatal activation during the MID task in anticipation of gain in the ITT population with and without imputation for missing data ($P<0.02$), in the per-protocol and as-treated populations ($P<0.02$) and in the completer population ($P<0.025$). We did not find a statistically significant site effect for the primary outcome measure.

Secondary outcomes. Prespecified analyses included group comparison on mean Snaith–Hamilton Pleasure Scale (SHAPS) score and change in Probabilistic Reward Task (PRT) response bias from block 1 to block 2 carried out in the ITT population.

SHAPS. A significant group effect emerged reflecting significantly lower baseline-adjusted mean SHAPS score after treatment with JNJ-67953964 in comparison to placebo ($F(1,86)=3.35$, $P=0.0345$; Hedges' $g=0.44$, 95% confidence interval = [0.012, 0.86]; Table 2 and Fig. 3a). Confirmatory support for this significant effect included that the between-group effect was statistically significant (1) in the per-protocol and as-treated populations ($P<0.025$) and (2) in the completer population ($P<0.05$). We found a significant site effect for the SHAPS score ($P<0.03$; Supplementary Table 3).

PRT. Seventy-six participants passed an a priori–defined quality-control check for at least baseline PRT data (performed with blinding to drug randomization; Supplementary Note) and were included in analyses. The groups did not differ in the rich-to-lean reward ratio ($F(1,53)=0.42$; $P=0.53$) or discriminability ($F(1,53)=5.72$; $P=0.20$), indicating that they were exposed to a similar reinforcement schedule and task difficulty, respectively. Contrary to our hypothesis, the block(1, 2)×treatment arm(JNJ-67953964, placebo)×time(baseline, treatment week 8) analysis of response bias was not significant ($F(1,53)=0.5$, $P=0.48$). However, the treatment arm×time interaction was significant ($F(1,53)=3.44$, $P=0.030$; Hedges' $g=0.49$, 95% confidence interval = [0.45, 0.54]), driven by higher post-treatment response bias with JNJ-67953964 relative to placebo (Table 2 and Fig. 3b).

Exploratory outcomes. Statistically significant group effects emerged for the mean and maximum fMRI ventral striatal activation during anticipation of loss as contrasted with no-incentive cue trials. Relative to placebo, JNJ-67953964 was associated with greater mean ($F(1,86)=11.7$, $P<0.001$; Hedges' $g=1.12$, 95% confidence interval = [0.96, 1.21]; Table 2) and maximum ($P<0.004$; Hedges' $g=0.66$, 95% confidence interval = [0.51, 0.85]) ventral striatal activation. These effects were also statistically significant in the per-protocol and as-treated (mean, $P<0.001$; maximum, $P<0.008$) and completer (mean, $P<0.001$; maximum, $P<0.012$) populations.

Statistically significantly greater enhancement of the consummatory subscale of the Temporal Experience of Pleasure Scale (TEPS) was also found with JNJ-67953964 versus placebo ($F(1,86)=4.3$,

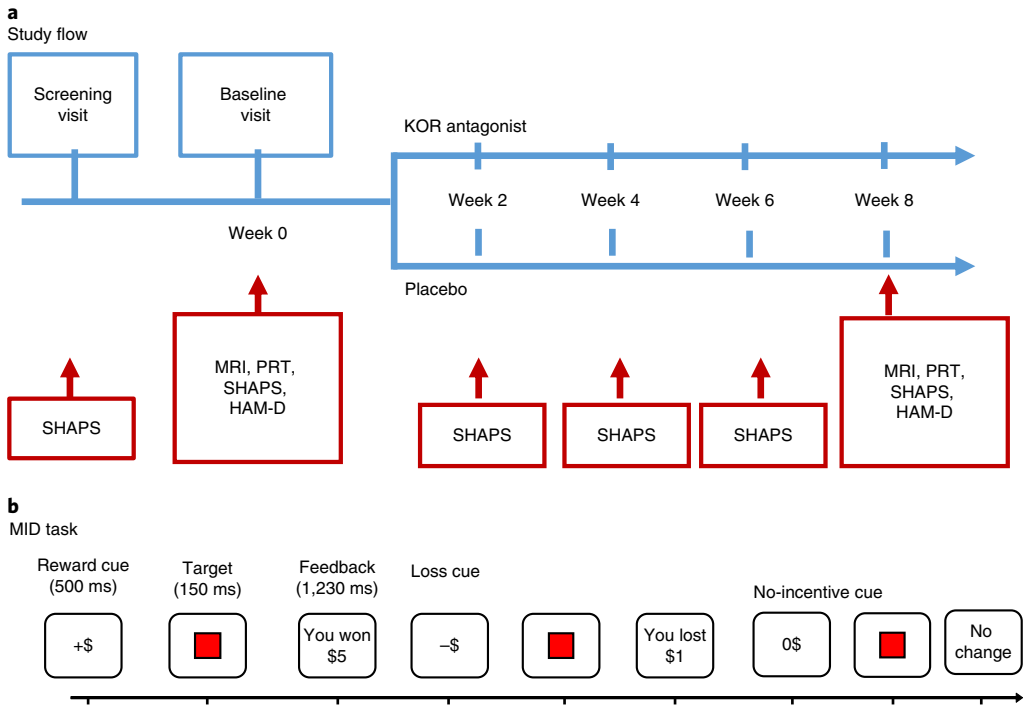


Fig. 1 | Study methods overview. a, Summary of study flow. Within 30 d of screening, participants returned for a baseline visit, which included administration of the Snaith–Hamilton Pleasure Scale (SHAPS)[44–47], which assesses anhedonic symptoms, Hamilton depression- and anxiety-rating scales[48,49], MRI, electroencephalogram (EEG) and the Probabilistic Reward Task (PRT)[35–41]. Participants were then randomized to JNJ-67953964 (10 mg) or placebo (1:1 ratio) for 8 weeks. After 8 weeks of double-blind treatment, participants underwent MRI, EEG, PRT, SHAPS and completed anxiety and depression scales and treatment was discontinued. At the final visit (week 12), participants were assessed for possible adverse effects. The primary outcome measure was ventral striatal activation during reward anticipation, assessed with fMRI during the Monetary Incentive Delay (MID) task[50–60]. Secondary measures were the SHAPS and response bias in the PRT. The PRT is a computerized task that objectively measures participants' ability to modulate behavior as a function of reinforcement history[35]. **b**, Trial structure of the MID task. During fMRI, participants performed four runs of the MID task (24 trials per run). For each trial (6 s), participants were presented with one of three possible cue shapes for 500 ms, which signaled whether the upcoming trial had the potential for monetary gain ($n=40$; denoted by +$), had the potential for monetary loss ($n=40$; denoted by –$) or was a no-incentive trial ($n=40$; denoted by 0$). Trial types were pseudorandomly ordered within each run. After 2,250–3,750 ms, a red square target was presented for 150 ms. Participants were instructed that, for reward trials, they could win money if they responded quickly to the target; for penalty trials, they could avoid losing money if they responded quickly to the target; and for no-incentive trials there would be no monetary change, but they should still respond quickly to the target. Participants received feedback 2,400–3,900 ms after target presentation. To standardize task difficulty, the 66th percentile of the reaction times collected during a practice session was used to determine wins versus penalties.

$P<0.02$; Hedges' $g=0.51$, 95% confidence interval = [–0.37, 1.41]; Table 2), but only in the ITT population.

No significant between-group differences were found for any of the other exploratory variables studied (Table 2).

Additional post hoc exploratory analyses. Exploratory analysis indicated that the degree of change in ventral striatal activation with treatment was significantly inversely correlated with baseline ventral striatal activation ($r=-0.49$, $P<0.0001$). This anticorrelation was somewhat larger in participants receiving JNJ-67953964 ($r=-0.60$, $P<0.0002$) than in those receiving placebo ($r=-0.50$, $P<0.001$). Logistic regression analysis indicated that baseline ventral striatal activation significantly predicted, at the individual level, which participants would be responders (defined on the basis of a median split of the change in mean ventral striatal activation) (Wald chi-squared = 12.4, $P<0.0004$; 76% accuracy in predicting responder status).

See Supplementary Table 3 for the results of exploratory analyses of site × time and site × treatment arm × time interactions for all variables.

Additionally, we assessed the relatedness of our primary outcome to the key secondary outcomes. We found that change in our primary outcome with treatment was significantly correlated with change in SHAPS score ($r=0.2$, $P<0.05$) and was associated with change in the block 1-to-block 2 difference in PRT response bias at a trend level ($r=0.23$, $P<0.071$).

Finally, we computed the treatment effect size for our primary outcome separately for the JNJ-67953964 and placebo groups. This measure increased from baseline with JNJ-67953964 with an effect size of 0.30 (Hedges' g) and decreased from baseline with placebo with an effect size of 0.52.

Safety. JNJ-67953964 was not associated with any serious adverse events and was generally well tolerated. No individuals discontinued participation owing to medication-related side effects other than worsening of depression or anxiety. Side effects occurring at least 5% more frequently with JNJ-67953964 than with placebo that were more than mild included pruritis (11.1%), depression exacerbation (6.7%) and rash (6.7%) (Table 3). There were no clinically meaningful changes in physical examination, vital signs, electrocardiogram (ECG) or laboratory tests.

Discussion

The present results establish that a dose of a medication documented to have robust KOR antagonism on the basis of PET imaging[24] has the hypothesized effect on neural function, thereby establishing

Table 1 | Study participant demographic and baseline data for the ITT population

Variable	JNJ-67953964 ($n=45$)	Placebo ($n=44$)	Total ($n=89$)
Mean age in years (s.d.)	40.7 (13.3)	38.2 (13.0)	39.5 (13.2)
Sex, % female	64.4	61.4	62.9
Race, %			
Caucasian	70.5	65.1	67.8
African American	22.7	18.6	20.7
Asian	2.3	4.7	3.4
American Indian/Alaskan Native	0.0	2.3	1.1
More than one race	4.5	9.3	6.9
Ethnicity, % Hispanic origin	11.6	11.6	11.6
Mean BMI (s.d.)	29.4 (6.4)	28.0 (5.9)	28.7 (6.2)
Mean weight in lbs (s.d.)	180.9 (43.7)	180.3 (40.6)	180.6 (41.9)
Mean baseline fMRI ventral striatal activation in MID task in anticipation of gain contrasted with no-incentive trials (s.d.)[a]	0.63 (0.9) ($n=44$)	0.64 (0.8) ($n=44$)	0.63 (0.8) ($n=88$)
Mean maximum baseline fMRI ventral striatal activation in MID task (s.d.) in anticipation of gain contrasted with no-incentive trials (s.d.)	2.66 (1.2) ($n=44$)	2.73 (1.2) ($n=44$)	2.70 (1.2) ($n=88$)
Mean baseline fMRI ventral striatal activation in MID task in anticipation of loss contrasted with no-incentive trials (s.d.)	0.29 (0.8) ($n=44$)	0.36 (0.7) ($n=44$)	0.33 (0.7) ($n=88$)
Mean maximum baseline fMRI ventral striatal activation in MID task (s.d.) in anticipation of loss contrasted with no-incentive trials (s.d.)	2.15 (1.2) ($n=44$)	2.23 (0.9) ($n=44$)	2.19 (1.0) ($n=88$)
Mean baseline PRT change in response bias from block 1 to block 2 (s.d.)[b]	0.02 (0.2) ($n=35$)	0.05 (0.2) ($n=41$)	0.04 (0.2) ($n=76$)
Mean baseline SHAPS (s.d.)[b]	36.4 (8.5) ($n=44$)	33.4 (5.9) ($n=44$)	34.9 (7.4) ($n=88$)
Mean baseline PRT response bias (averaged across blocks) (s.d.)	0.108 (0.027) ($n=35$)	0.113 (0.025) ($n=41$)	0.111 (0.026) ($n=76$)
Mean baseline EEfRT (s.d.)	0.35 (0.2) ($n=42$)	0.38 (0.2) ($n=41$)	0.36 (0.2) ($n=83$)
Mean baseline TEPS anticipatory subscore (s.d.)	29.3 (5.7) ($n=44$)	29.5 (5.6) ($n=44$)	29.4 (5.7) ($n=88$)
Mean baseline TEPS consummatory subscore (s.d.)	26.3 (4.4) ($n=44$)	26.1 (4.4) ($n=44$)	26.2 (4.4) ($n=88$)
Mean baseline VAS anhedonia (s.d.)	2.93 (2.1) ($n=44$)	3.59 (2.2) ($n=44$)	3.26 (2.2) ($n=88$)
Mean baseline resting-state EEG delta current density in rostral anterior cingulate (s.d.)	73.0 (97.1) ($n=43$)	75.2 (51.6) ($n=38$)	74.0 (78.6) ($n=81$)
Mean baseline HAM-D (s.d.)	16.3 (5.2)	14.8 (5.9)	15.6 (5.6)
Mean baseline HAM-A (s.d.)	16.0 (5.8)	15.1 (6.6)	15.5 (6.2)
Mean baseline CGI-S (s.d.)	3.9 (0.6)	4.0 (0.5)	3.9 (0.5)
Mean baseline CPFQ (s.d.)	27.2 (6.4) ($n=44$)	25.4 (5.7) ($n=44$)	26.3 (6.1) ($n=44$)

[a]A priori–specified primary outcome measure. [b]A priori–specified secondary outcome measure. No variables were different between groups at a significance level of $P<0.05$. There were 45 participants in the JNJ-67953964 group and 44 participants in the placebo group unless otherwise noted; when n values were lower, it was because of missing baseline data for that variable. BMI, body-mass index; EEfRT, Effort Expenditure for Rewards Task; TEPS, Temporal Experience of Pleasure Scale; VAS, Visual Analog Scale; HAM-D, Hamilton Depression-Rating Scale; HAM-A, Hamilton Anxiety-Rating Scale; CGI-S, Clinical Global Impression of Severity; CPFQ, Cognitive and Physical Functioning Questionnaire.

the POM that engaging this target has an effect on a brain function implicated in hedonic responses. This, in turn, suggests that JNJ-67953964 has potential as a specific anhedonia therapy.

The strength of the conclusion is reinforced by the fact that a significant JNJ-67953964 versus placebo effect was evident in all measures assessing ventral striatal activity. In addition to the statistically significant effect observed with the primary outcome measure, we found significant effects in comparison to placebo on maximum ventral striatal activation in anticipation of gain and mean and maximum activation in anticipation of loss.

Notably, while there was an increase in the primary outcome measure in the JNJ-67953964 group with treatment, a substantial amount of the JNJ-67953964 versus placebo difference was driven by a decrease over time in the placebo group. This impression was verified in post hoc exploratory analysis. However, while the decrease from baseline in the placebo group (Hedges' $g=0.52$) was larger than the increase in the JNJ-67953964 group, the effect size for the increase with treatment in the JNJ-6795364 group was nontrivial in magnitude (Hedges' $g=0.30$). Our findings are consistent with previous evidence of diminished ventral striatal response to anticipation of reward with repeated testing. For example, the degree of activation in anticipation of reward in the MID task decreased with repeated testing in the same session in 26 of 29 individuals with anhedonia[29]. Furthermore, healthy control individuals experienced a decrease (not statistically significant) in ventral striatal activation in anticipation of reward across two test sessions separated by an average of 48 d (ref. [30]). To mitigate the possible effects of such adaptation, our primary analysis involved a direct comparison between the JNJ-67953964 and placebo groups. When viewed in the context of previous findings[29,30], our results imply that JNJ-67953964 led to an increase in ventral striatal activation with an effect size of 0.30 in spite of possible habituation effects stemming from repeated administration (as seen in the placebo group).

Whatever the source of the decrease in striatal activation in anticipation of reward in participants administered placebo, on the basis of the a priori–planned approach of assessing statistical

Table 2 | Efficacy results based on mixed-effects models in the ITT population represented as baseline-corrected mean (s.d.) at the end of the double-blind treatment period

Variable	JNJ-67953964	Placebo	*n*	*P*	Effect size (Hedges' *g*)
Primary outcome measure					
Mean fMRI ventral striatal activation in MID task in anticipation of gain contrasted with no-incentive trials	0.72 (0.67)	0.33 (0.68)	88	0.0095	0.58
Secondary outcome measures					
Mean SHAPS	30.8 (3.7)	32.4 (3.6)	88	0.0345	0.44
PRT change in response bias from block 1 to block 2	0.059 (0.15)	0.066 (0.15)	76	>0.10	0.10
Exploratory outcome measures					
Maximum fMRI ventral striatal activation in MID task in anticipation of gain contrasted with no-incentive trials	2.84 (0.86)	2.36 (0.86)	88	0.012	0.55
Mean fMRI ventral striatal activation in MID task in anticipation of loss contrasted with no-incentive trials	0.73 (0.6)	0.07 (0.6)	88	<0.001	1.12
Maximum fMRI ventral striatal activation in MID task in anticipation of loss contrasted with no-incentive trials	2.73 (0.8)	2.18 (0.8)	88	0.0035	0.66
Mean PRT response bias	0.153 (0.013)	0.070 (0.018)	76	0.030	0.49
TEPS anticipatory subscale	32.8 (5.5)	32.6 (5.4)	88	>0.10	0.03
TEPS consummatory subscale	29.3 (4.2)	27.1 (4.2)	88	0.017	0.51
EEfRT	0.41 (0.11)	0.41 (0.11)	83	>0.10	0.00
VAS anhedonia	4.2 (1.5)	4.6 (1.5)	88	>0.10	0.30
Resting-state EEG delta current density in rostral anterior cingulate	62.0 (77.9)	81.2 (81.1)	81	>0.10	0.24
HAM-D	10.8 (4.0)	11.1 (3.9)	89	>0.10	0.09
HAM-A	11.0 (4.2)	10.6 (4.4)	89	>0.10	0.09
CGI-I	3.27 (0.7)	3.20 (0.6)	89	>0.10	0.08
CGI-S	3.28 (0.5)	3.32 (0.5)	89	>0.10	0.08
CPFQ	21.1 (4.2)	21.1 (4.2)	88	>0.10	0.00

Means are baseline-corrected values derived from mixed-effects models. CGI-I, Clinical Global Impression of Improvement.

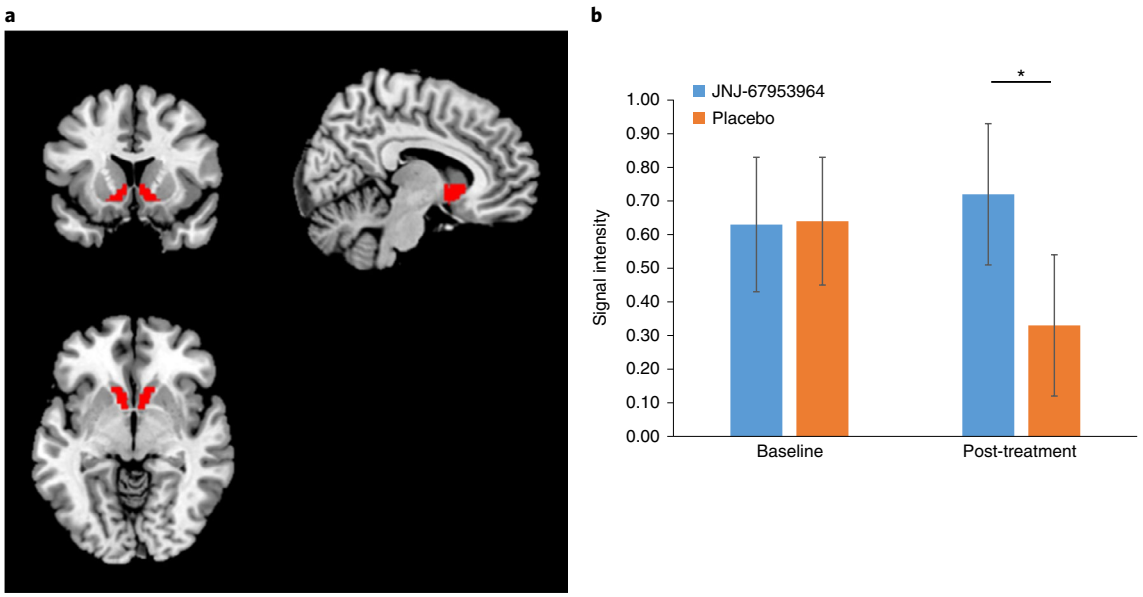


Fig. 2 | Results for the primary outcome measure: mean fMRI ventral striatal activation in anticipation of rewards in the monetary incentive delay task. a, Location of the ventral striatal region of interest (ROI) based on the Harvard–Oxford Subcortical Atlas. **b**, Mean signal intensity during reward anticipation within the ventral striatal ROI before and after treatment with JNJ-67953964 or placebo. Error bars represent the 95% confidence intervals around the mean signal intensity. *$P < 0.01$. Mixed-effects model analysis was carried out in the ITT population including mean centered baseline value, age, study site and sex as covariates.

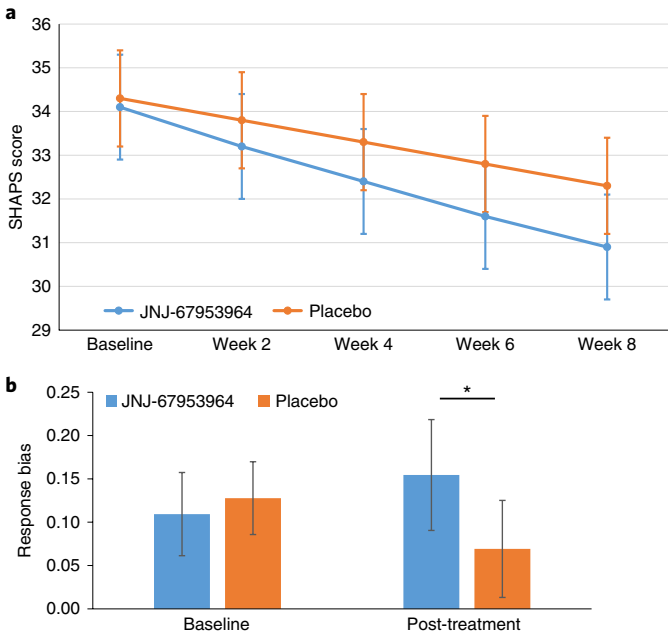


Fig. 3 | SHAPS and PRT results. a, Effects of study drug versus placebo on mean SHAPS score (ITT population). Error bars represent the 95% confidence intervals around the means. **b**, Effects of study drug versus placebo on mean PRT response bias (ITT population). Error bars represent standard errors. *$P < 0.05$. Mixed-effects model analysis was carried out in the ITT population including mean centered baseline value, age, study site and sex as covariates.

significance in a JNJ-67953964 versus placebo comparison, this study robustly establishes POM for KOR antagonism. This provides a rational basis for proceeding with development of this target and carrying out larger trials powered for the use of clinical endpoints to determine the clinical impact of engaging this target with JNJ-67953964.

According to the fast-fail framework, establishing POM for KOR antagonism achieves two critical objectives for future studies with clinical endpoints:

1) It increases the likelihood that positive outcomes reflect drug effects on the brain circuitry hypothesized to mediate therapeutic effects, thereby decreasing the likelihood that they reflect nonspecific effects or bias to which such studies are relatively vulnerable;

2) It ensures that negative results are interpretable, indicating that achieving the hypothesized effect on brain function does not lead to the expected clinical effect. Otherwise, such negative outcomes could reflect insufficient dosage to engage hypothesized circuitry or that the phase 2 findings were false positives reflecting nonspecific effects or bias.

It is critical to appreciate the priority placed on our neuroimaging-based primary outcome in determining the promise of KOR antagonism. The fast-fail approach focuses on biomarker-based outcomes in phase 2 to assess, as directly as possible, the circuitry hypothesized to mediate treatment effects. Biomarkers, being closer to the direct biological effects of the drug than clinical measures, are assumed to be associated with effect sizes that are sufficiently large to be detected with studies of the size typically carried out in phase 2. According to our model, effects seen on clinical symptom outcomes are unreliable with regard to providing information on whether the drug has the hypothesized brain effect. Having a significant effect on the primary biomarker-based outcome is sufficient for proceeding to trials with symptom measures. Significant effects on all of the symptom measures (SHAPS, TEPS and VAS) but not on the primary neuroimaging-based outcome would not support

Table 3 | Adverse effects of more than mild severity

Side effect	JNJ-67953964 ($n = 45$)		Placebo ($n = 44$)	
	n[a]	%[b]	n[a]	%[b]
Headache	5	11.1	4	9.1
Pruritus	5	11.1	1	2.3
Anxiety	3	6.7	2	4.5
Insomnia	3	6.7	2	4.5
Suicidal ideation	3	6.7	2	4.5
Diarrhea	1	2.2	3	6.8
Constipation	2	4.4	1	2.3
Depression	3	6.7	0	0.0
Dizziness	2	4.4	1	2.3
Fatigue	1	2.2	2	4.5
Rash	3	6.7	0	0.0
Restlessness	1	2.2	2	4.5
Vision blurred	2	4.4	1	2.3
Arthralgia	1	2.2	1	2.3
Disturbance in attention	1	2.2	1	2.3
Dry mouth	1	2.2	1	2.3
Nausea	0	0.0	2	4.5
Panic attack	1	2.2	1	2.3
Pollakiuria	2	4.4	0	0.0
Pyrexia	0	0.0	2	4.5
Back pain	0	0.0	1	2.3
Blepharitis	1	2.2	0	0.0
Chest pain	1	2.2	0	0.0
Coordination abnormal	0	0.0	1	2.3
Costochondritis	1	2.2	0	0.0
Cough	0	0.0	1	2.3
Dry skin	0	0.0	1	2.3
Dysuria	1	2.2	0	0.0
Eye pruritus	0	0.0	1	2.3
Fall	0	0.0	1	2.3
Hyperhidrosis	0	0.0	1	2.3
Hypersomnia	0	0.0	1	2.3
Irritability	1	2.2	0	0.0
Libido decreased	0	0.0	1	2.3
Muscle twitching	1	2.2	0	0.0
Musculoskeletal chest pain	0	0.0	1	2.3
Nasopharyngitis	0	0.0	1	2.3
Night sweats	0	0.0	1	2.3
Palpitations	0	0.0	1	2.3
Self-injurious ideation	1	2.2	0	0.0
Tendon rupture	1	2.2	0	0.0
Toothache	0	0.0	1	2.3
Upper respiratory tract infection	0	0.0	1	2.3
Viral infection	1	2.2	0	0.0

[a]The number of participants experiencing the event at least once (not the total number of events).
[b]The percentage of participants who experienced the event.

proceeding. At the same time, having supportive significant effects on symptom measures is of value and increases confidence in the likelihood of success of development. However, the presence of such effects is not a precondition for meeting 'go' criteria for drug development because such effects are likely to be too small to be detected in this type of study.

It is also important to note that we cannot cite examples that demonstrate the effectiveness of this fast-fail framework. This effort is the first to ever attempt implementing the fast-fail approach, which speaks to the high significance of our study and sets the stage for the first time for carrying out the phase 3 studies needed to determine the utility of fast-fail approaches.

The methods of this study also serve as a model for implementing fast-fail early-phase methodology and specifically establish fast-fail methods for the development of anhedonia treatments. It should be noted that, in carrying out fast-fail POM studies, failure to establish POM may not indicate that engaging the target does not impact the relevant neural circuitry. It could also reflect insufficient sensitivity of the outcome measure. This is a key challenge for fast-fail approaches given the limited number of POM-study-suitable outcome measures with well-characterized statistical properties. In this regard, our effort establishes the usefulness for POM studies of the MID fMRI measure. Lastly, our findings generally support the feasibility of implementing the fast-fail approach.

This study brings us closer to our original vision of how fast-fail methods should be implemented in research studies and ultimately applied in clinical practice. This vision included that biomarkers would be used to select and effectively phenotype patients. We originally intended to select participants on the basis of ventral striatal activation, but it was felt that insufficient work had been completed characterizing the statistical properties of our fMRI measure to allow us to do so. However, the results of our exploratory analysis indicating that baseline mean ventral striatal activation in anticipation of gain was a significant predictor of both the change in this variable with treatment and responder status suggest the promise of using fMRI mean ventral striatal activation in anticipation of reward for selection of participants who are likely to respond to KOR antagonist therapy. As such, we anticipate that this study and additional studies carried out with measures of ventral striatal activation in appropriately selected participant populations will make it possible to select participants by using these measures for fast-fail studies of potential therapies for anhedonia and, ultimately, for optimizing clinical practice.

Evidence that the effects on neural circuitry were accompanied by effects on clinical measures of anhedonia further reinforces the promise of KOR antagonism as an anhedonia therapy. Statistically significantly greater improvement was found with JNJ-67953964 than with placebo on the SHAPS score (an a priori–specified secondary outcome) as well as on the consummatory subscale of TEPS (an exploratory outcome). At the same time, there was some divergence in outcomes in terms of a lack of significant effects on some exploratory measures of reward-related function (VAS anhedonia, EEfRT and TEPS anticipatory subscale). Although the divergence in outcomes could indicate limited replicability of the effect on the primary outcome measure with clinical and behavioral measures, another possible explanation is that our study was underpowered to detect effects on these measures, which were hypothesized to be smaller than effects on the neuroimaging measure for which we estimated power and sample size. Another possibility is that our reward-related measures might assess various aspects of reward-related function that are differentially impacted by KOR antagonism. Collectively, these findings suggest that the observed brain effect may mediate clinically meaningful anhedonic effects but that further work is needed to determine the relationship between such brain effects and clinical and behavioral measures of anhedonia.

Although there was a statistically significant effect on the key secondary clinical outcome measure of the SHAPS score, the size of this effect is of uncertain clinical importance. We are aware of no accepted, standard means for determining clinical importance or clinical effect size with the SHAPS. However, the treatment effect size (Hedges' $g=0.44$) would be considered just smaller than medium sized when compared against the most commonly used benchmarks (small, 0.2; medium, 0.5; large, 0.8)[31].

In terms of the behavioral measures, we did not find a significant effect in the preplanned behavioral secondary analysis (treatment arm × block × time interaction for PRT response bias)[32,33]. However, we did find a significant treatment arm × time interaction, which also implies that the groups differed in how strongly their behavior was modulated by rewards. Factors related to why we did not find a significant effect for the planned PRT analysis are reviewed in the Supplementary Note.

An important consideration with respect to the PRT analyses is that approximately 17% of the participants failed the a priori–defined quality-control evaluation for at least one of their two PRT assessments. This study is among the first to implement the PRT in a multisite randomized controlled trial. Although standardization and training across sites was implemented, it was evidently challenging to maintain high reliability across sites and across years of a complex randomized controlled trial. This rate of data loss is similar to what emerged in another large multisite randomized controlled trial (EMBARC)[34], but was substantially greater than previous studies using the PRT[35–41]. It is clear that the significant data loss for the PRT decreased the available power for PRT analyses and was a limitation. Accordingly, further work is needed to limit data loss associated with this behavioral measure in the context of multisite studies.

Nonetheless, it was the case that we found significant effects for JNJ-67953964 across all three units of analysis—brain circuitry, behavior and self-report. Notably, the results of exploratory analysis indicating that the change with treatment in our primary outcome (fMRI measure) was statistically significantly correlated with the change in a self-report measure (the SHAPS score) and was associated at a trend level with our key behavioral secondary outcome (the PRT) supports the idea that KOR antagonism had a coherent effect on measures of anhedonia across units of analysis and increases confidence in the likelihood of success of development.

A comparison of the relative treatment effect sizes seen across various measures of anhedonia supports a key principle of the fast-fail approach, namely that biomarkers should be associated with greater effect sizes when used as treatment outcome measures than clinical and behavioral measures. Specifically, we found that the effect size associated with the primary neuroimaging measure (Hedges' $g=0.58$) was greater than the effect size associated with a behavioral measure (Hedges' $g=0.49$) and the key secondary self-report measure (Hedges' $g=0.44$). As a result, this study provides an example supporting the notion that the use of biomarkers in early-phase studies could increase power, thereby allowing smaller studies, which are more likely to be reliable and replicated.

It is notable that the largest drug versus placebo treatment effect size was seen with ventral striatal activity in anticipation of loss. While this may seem inconsistent with an intervention that improves anhedonia, it is in line with previous work. Several MID studies have reported robust fMRI ventral striatal activation during anticipation of losses or aversive events, and this activation correlated with ventral striatal activation occurring during anticipation of gains[30,42,43]. These observations led to the hypothesis that ventral striatal activation reflects the 'motivational relevance of an upcoming event' and is not specific to anticipation of reward[42]. Taken in this context, the findings suggest the hypothesis that anhedonia may be best thought of as loss of motivational relevance rather than impairment specific to reward function.

Another notable finding of this study is that KOR antagonism led to improvement in measures of anhedonia but did not improve broad measures of depression (HAM-D) or anxiety (HAM-A). This suggests that, although anhedonia is a core feature of MDD, it is possible to distinguish effects on anhedonia from antidepressant effects. A limitation to this conclusion, however, is that this study was not designed to nor capable of rigorously determining the effects of KOR antagonism on depression. There was no minimum depression severity required for participation, nor were individuals required to have MDD to participate (20% had an anxiety disorder or post-traumatic stress disorder (PTSD) and not MDD). The initial depression and anxiety severity in this study were not high, decreasing the possibility for improvement in HAM-D and HAM-A and likely decreasing treatment effect sizes. Also, the anhedonia factor for HAM-D only corresponds to several of the items from this scale, diminishing the likelihood that a treatment effect would be found with this measure due to a therapy that specifically improves anhedonia. Nonetheless, the effect sizes seen for anhedonia measures are so much greater than the effect sizes for depression and anxiety that it seems reasonable to conclude that KOR antagonism likely has a specific therapeutic effect on anhedonia and not a broad effect on depression and anxiety and that it is possible to distinguish anhedonic effects from more general antidepressant/anxiolytic effects. As such, this study opens the door to developing treatments specifically targeting anhedonia.

Lastly, this study had a number of limitations not mentioned above. We did not ask participants whether they believed they had received study medication or placebo. As a result, we could not assess whether participants were unblinded. We also had limited ability to examine interaction effects. For example, because of the number of covariates and sites relative to the sample size, we did not plan to examine site × time and site × treatment arm × time interactions (post hoc analysis results appear in Supplementary Table 3). Further, we had a dropout rate of approximately 25%, slightly higher than anticipated (20%), which limited the power available for completer analyses.

Still, this study is notable because it represents the first successful implementation of the NIMH fast-fail treatment development approach. It is hoped that it will set the stage for future applications of the fast-fail approach, which have the potential to lead to more efficient treatment development, thereby facilitating an increase in investment in developing much-needed psychiatric therapies.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41591-020-0806-7.

Received: 5 May 2019; Accepted: 20 February 2020;
Published online: 30 March 2020

References

1. Denys, D. & de Geus, F. Predictors of pharmacotherapy response in anxiety disorders. *Curr. Psychiatry Rep.* **7**, 252–257 (2005).
2. Rush, A. J. et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. *Am. J. Psychiatry* **163**, 1905–1917 (2006).
3. Insel, T. R. & Wang, P. S. The STAR*D trial: revealing the need for better treatments. *Psychiatr. Serv.* **60**, 1466–1467 (2009).
4. National Advisory Mental Health Workgroup. *From Discovery to Cure: Accelerating the Development of New and Personalized Interventions for Mental Illness* (NIMH, 2010).
5. Paul, S. M. et al. How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat. Rev. Drug Discov.* **9**, 203–214 (2010).
6. Kesselheim, A. S., Hwang, T. J. & Franklin, J. M. Two decades of new drug development for central nervous system disorders. *Nat. Rev. Drug Discov.* **14**, 815–816 (2015).
7. Brady, L. S. & Insel, T. R. Translating discoveries into medicine: psychiatric drug development in 2011. *Neuropsychopharmacology* **37**, 281–283 (2012).
8. Krystal, A. D. et al. The first implementation of the NIMH FAST-FAIL approach to psychiatric drug development. *Nat. Rev. Drug Discov.* **18**, 82–84 (2018).
9. Wiedemann, K. Biomarkers in development of psychotropic drugs. *Dialogues Clin. Neurosci.* **13**, 225–234 (2011).
10. Borsook, D., Becerra, L. & Fava, M. Use of functional imaging across clinical phases in CNS drug development. *Transl. Psychiatry* **16**, e282 (2013).
11. Insel, T. et al. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am. J. Psychiatry* **167**, 748–751 (2010).
12. Whitton, A. E., Treadway, M. T. & Pizzagalli, D. A. Reward processing dysfunction in major depression, bipolar disorder and schizophrenia. *Curr. Opin. Psychiatry* **28**, 7–12 (2015).
13. Carlezon, W. A. Jr & Krystal, A. D. κ-Opioid antagonists for psychiatric disorders: from bench to clinical trials. *Depress. Anxiety* **33**, 895–906 (2016).
14. Carlezon, W. A. Jr et al. Depressive-like effects of the κ-opioid receptor agonist salvinorin A on behavior and neurochemistry in rats. *J. Pharmacol. Exp. Ther.* **316**, 440–447 (2006).
15. Bruijnzeel, A. W. κ-Opioid receptor signaling and brain reward function. *Brain Res. Rev.* **62**, 127–146 (2009).
16. Chartoff, E. et al. Blockade of κ-opioid receptors attenuates the development of depressive-like behaviors induced by cocaine withdrawal in rats. *Neuropharmacology* **62**, 167–176 (2012).
17. Ebner, S. R., Roitman, M. F., Potter, D. N., Rachlin, A. B. & Chartoff, E. H. Depressive-like effects of the κ-opioid receptor agonist salvinorin A are associated with decreased phasic dopamine release in the nucleus accumbens. *Psychopharmacology* **210**, 241–252 (2010).
18. Maisonneuve, I. M., Archer, S. & Glick, S. D. U50,488, a κ-opioid receptor agonist, attenuates cocaine-induced increases in extracellular dopamine in the nucleus accumbens of rats. *Neurosci. Lett.* **181**, 57–60 (1994).
19. Muschamp, J. W. et al. Activation of CREB in the nucleus accumbens shell produces anhedonia and resistance to extinction of fear in rats. *J. Neurosci.* **31**, 3095–3103 (2011).
20. Tomasiewicz, H. C., Todtenkopf, M. S., Chartoff, E. H., Cohen, B. M. & Carlezon, W. A. Jr. The κ-opioid agonist U69,593 blocks cocaine-induced enhancement of brain stimulation reward. *Biol. Psychiatry* **64**, 982–988 (2008).
21. Wee, S. & Koob, G. F. The role of the dynorphin–kappa opioid system in the reinforcing effects of drugs of abuse. *Psychopharmacology* **210**, 121–135 (2010).
22. Rorick-Kehn, L. M. et al. LY2456302 is a novel, potent, orally-bioavailable small molecule κ-selective antagonist with activity in animal models predictive of efficacy in mood and addictive disorders. *Neuropharmacology* **77C**, 131–144 (2013).
23. Lowe, S. L. et al. Safety, tolerability, and pharmacokinetic evaluation of single- and multiple-ascending doses of a novel κ-opioid receptor antagonist LY2456302 and drug interaction with ethanol in healthy subjects. *J. Clin. Pharmacol.* **54**, 968–978 (2014).
24. Zheng, M. Q. et al. Synthesis and evaluation of 11C-LY2795050 as a κ-opioid receptor antagonist radiotracer for PET imaging. *J. Nucl. Med.* **54**, 455–463 (2013).
25. Carlezon, W. A. Jr, Béguin, C., Knoll, A. T. & Cohen, B. M. κ-Opioid ligands in the study and treatment of mood disorders. *Pharmacol. Ther.* **123**, 334–343 (2009).
26. Knutson, B. & Gibbs, S. E. Linking nucleus accumbens dopamine and blood oxygenation. *Psychopharmacology* **191**, 813–822 (2007).
27. Schott, B. H. et al. Mesolimbic functional magnetic resonance imaging activations during reward anticipation correlate with reward-related ventral striatal dopamine release. *J. Neurosci.* **28**, 14311–14319 (2008).
28. Hedges, L. Distribution theory for Glass's estimator of effect size and related estimators. *J. Educat. Stat.* **6**, 107–128 (1981).
29. Green, I. W., Pizzagalli, D. A., Admon, R. & Kumar, P. Anhedonia modulates the effects of positive mood induction on reward-related brain activation. *Neuroimage* **193**, 115–125 (2019).
30. Stoy, M. et al. Hyporeactivity of ventral striatum towards incentive stimuli in unmedicated depressed patients normalizes after treatment with escitalopram. *J. Psychopharmacol.* **26**, 677–688 (2012).
31. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences* (Routledge, 1988).
32. Vrieze, E. et al. Reduced reward learning predicts outcome in major depressive disorder. *Biol. Psychiatry* **73**, 639–645 (2013).
33. Kaiser, R. H. et al. Frontostriatal and dopamine markers of individual differences in reinforcement learning: a multi-modal investigation. *Cereb. Cortex* **28**, 4281–4290 (2018).

34. Webb, C. A. et al. Neural correlates of three promising endophenotypes of depression: evidence from the EMBARC study. *Neuropsychopharmacology* **41**, 454–463 (2016).
35. Pizzagalli, D. A., Jahn, A. L. & O'Shea, J. P. Toward an objective characterization of an anhedonic phenotype: a signal-detection approach. *Biol. Psychiatry* **57**, 319–327 (2005).
36. Pizzagalli, D. A. et al. Single dose of a dopamine agonist impairs reinforcement learning in humans: behavioral evidence from a laboratory-based measure of reward responsiveness. *Psychopharmacology* **196**, 221–232 (2008).
37. Pizzagalli, D. A., Iosifescu, D., Hallett, L. A., Ratner, K. G. & Fava, M. Reduced hedonic capacity in major depressive disorder: evidence from a probabilistic reward task. *J. Psychiatr. Res.* **43**, 76–87 (2008).
38. Santesso, D. L. et al. Individual differences in reinforcement learning: behavioral, electrophysiological, and neuroimaging correlates. *Neuroimage* **42**, 807–816 (2008).
39. Liu, W. et al. Deficits in sustaining reward responses in subsyndromal and syndromal major depression. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **35**, 1045–1052 (2011).
40. Fletcher, K. et al. Anhedonia in melancholic and non-melancholic depressive disorders. *J. Affect. Disord.* **184**, 81–88 (2015).
41. Morris, B. H., Bylsma, L. M., Yaroslavsky, I., Kovacs, M. & Rottenberg, J. Reward learning in pediatric depression and anxiety: preliminary findings in a high-risk sample. *Depress. Anxiety* **32**, 373–381 (2015).
42. Jensen, J. et al. Direct activation of the ventral striatum in anticipation of aversive stimuli. *Neuron* **40**, 1251–1257 (2003).
43. Carter, R. M., McInnes, J. J., Huettel, S. A. & Adcock, R. A. Activation in the VTA and nucleus accumbens increases in anticipation of both gains and losses. *Front. Behav. Neurosci.* **3**, 1–15 (2009).
44. Snaith, R. P. et al. A scale for the assessment of hedonic tone the Snaith–Hamilton Pleasure Scale. *Br. J. Psychiatry* **167**, 99–103 (1995).
45. Franken, I. H., Rassin, E. & Muris, P. The assessment of anhedonia in clinical and non-clinical populations: further validation of the Snaith–Hamilton Pleasure Scale (SHAPS). *J. Affect. Disord.* **99**, 83–89 (2007).
46. Nakonezny, P. A. et al. Evaluation of anhedonia with the Snaith–Hamilton Pleasure Scale (SHAPS) in adult outpatients with major depressive disorder. *J. Psychiatr. Res.* **65**, 124–130 (2015).
47. Nakonezny, P. A., Carmody, T. J., Morris, D. W., Kurian, B. T. & Trivedi, M. H. Psychometric evaluation of the Snaith–Hamilton pleasure scale in adult outpatients with major depressive disorder. *Int. Clin. Psychopharmacol.* **25**, 328–333 (2010).
48. Hamilton, M. Development of a rating scale for primary depressive illness. *Br. J. Soc. Psychol.* **6**, 278–296 (1967).
49. Maier, W., Buller, R., Philipp, M. & Heuser, I. The Hamilton Anxiety Scale: reliability, validity and sensitivity to change in anxiety and depressive disorders. *J. Affect. Disord.* **14**, 61–68 (1988).
50. Knutson, B., Westdorp, A., Kaiser, E. & Hommer, D. FMRI visualization of brain activity during a monetary incentive delay task. *Neuroimage* **12**, 20–27 (2000).
51. Knutson, B., Bhanji, J. P., Cooney, R. E., Atlas, L. Y. & Gotlib, I. H. Neural responses to monetary incentives in major depression. *Biol. Psychiatry* **63**, 686–692 (2008).
52. Knutson, B., Fong, G. W., Adams, C. M., Varner, J. L. & Hommer, D. Dissociation of reward anticipation and outcome with event-related fMRI. *Neuroreport.* **12**, 3683–3687 (2001).
53. Knutson, B., Adams, C. M., Fong, G. W. & Hommer, D. Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *J. Neurosci.* **21**, RC159 (2001).
54. Knutson, B., Wimmer, G. E., Kuhnen, C. M. & Winkielman, P. Nucleus accumbens activation mediates the influence of reward cues on financial risk taking. *Neuroreport.* **19**, 509–513 (2008).
55. Carlson, J. M., Foti, D., Mujica-Parodi, L. R., Harmon-Jones, E. & Hajcak, G. Ventral striatal and medial prefrontal BOLD activation is correlated with reward-related electrocortical activity: a combined ERP and fMRI study. *Neuroimage* **57**, 1608–1616 (2011).
56. Dillon, D. G. et al. Dissociation of neural regions associated with anticipatory versus consummatory phases of incentive processing. *Psychophysiology* **45**, 36–49 (2008).
57. Keedwell, P. A., Andrew, C., Williams, S. C., Brammer, M. J. & Phillips, M. L. The neural correlates of anhedonia in major depressive disorder. *Biol. Psychiatry* **58**, 843–853 (2005).
58. Keller, J. et al. Trait anhedonia is associated with reduced reactivity and connectivity of mesolimbic and paralimbic reward pathways. *J. Psychiatr. Res.* **47**, 1319–1328 (2013).
59. Li, Z. et al. Improving motivation through real-time fMRI-based self-regulation of the nucleus accumbens. *Neuropsychology* **32**, 764–776 (2018).
60. Oldham, S. et al. The anticipation and outcome phases of reward and loss processing: a neuroimaging meta-analysis of the monetary incentive delay task. *Hum. Brain Mapp.* **39**, 3398–3418 (2018).

Methods

Detailed information regarding study methods appears in the accompanying Life Sciences Reporting Summary.

Patient population. Patients were eligible for enrollment if they were 21 to 65 years of age and had clinically significant anhedonia as defined by a SHAPS[44] score of at least 20 (as assessed by using dimensional scoring guidelines)[45]. The use of this cutoff was based on receiver-operating-characteristic curve analysis carried out with the SHAPS score for discriminating a Montgomery–Åsberg Depression Rating Scale (MADRS) anhedonia item score of greater than 4/6 (considered clinically significant) versus 4 or less[44]. In addition, participants had to currently meet the DSM-IV TR diagnostic criteria for MDD, bipolar I or II depressed, generalized anxiety disorder, social phobia, panic disorder or PTSD on the basis of the Mini-International Neuropsychiatric Interview for DSM-IV (MINI)[61,62].

We sought to include a cross-diagnostic sample of patients with anhedonia to be consistent with the NIMH RDoC dimensional approach to classification of mental disorders[8]. Because anhedonia is a core symptom of MDD and one of its diagnostic criteria, we were concerned that without constraining enrollment we might exclusively find and enroll individuals with anhedonia occurring in the setting of MDD. Lacking previous studies attempting to enroll a similar population as a guide, we estimated that it would be feasible to recruit at least 33% of the individuals who had an anxiety disorder without current MDD based on the MINI. Over the course of the study, we found recruiting individuals with anxiety disorders without current MDD more challenging than initially anticipated; as a result, with approval of the NIMH Data Safety Monitoring Board (DSMB) and institutional review boards, we decreased the target to 20%.

Participants underwent assessment of medical and psychiatric history, physical examination, laboratory testing and MINI assessment and, on this basis, were excluded if they were expected to require any hospitalization during the course of the study; had a history of a psychotic disorder, current manic or mixed episode, autism spectrum disorders or mental retardation; met DSM-IV TR criteria for substance abuse within the last 3 months or substance dependence within the last 6 months; had a history of an unstable or untreated serious medical condition; had active suicidal intent or plan or a history of an attempt within the past 3 months based on physician evaluation and the Columbia Suicide Severity Rating Scale (C-SSRS)[63]; used any medication with significant CNS effects, including antidepressants, antipsychotics, anxiolytics, anticonvulsants, mood-stabilizing agents, muscle relaxants, centrally acting antihistaminergics, stimulants or insomnia medications within five half-lives of baseline or at any time during the study; used any medication that is primarily metabolized by cytochrome P450 2C8 within 14 d of baseline or at any time during the study; had any contraindications to the MRI procedures; had a positive urine drug screen at any time during the study; used any investigational medication within 3 months of the study; had a history of gastric disease (including peptic ulcer disease, gastritis, upper gastrointestinal bleeding or any gastrointestinal precancerous condition) or had current clinically evident gastrointestinal complaints; had a positive urea breath test (exclusionary in the first half of the study, after which approval was obtained from the US Food and Drug Administration (FDA), the NIMH DSMB and the relevant institutional review boards to drop this requirement); had current use of a proton pump inhibitor or histamine-2 blocker or a history of chronic non-steroidal anti-inflammatory drug use; had a history of use of *Salvia divinorum* or used *S. divinorum* at any time during the study; smoked cigarettes or used other nicotine-containing products within the last month or at any time during the study; or were pregnant or lactating.

Trial design. The trial was conducted at six centers in the United States (Duke University, Yale University, Icahn School of Medicine at Mount Sinai, Baylor College of Medicine, Indiana University and Case Western Reserve University). The first patient was enrolled on 17 August 2015, and the last patient was enrolled on 29 August 2017. The study consisted of a randomized, double-blind, placebo-controlled, parallel-group, 8-week trial. Participants were randomized to JNJ-67953964 (10 mg) or placebo in a 1:1 ratio administered as identical-appearing tablets. Qualifying participants were provided the next container among sequentially numbered containers at each site through which the randomization sequence, established by the coordinating center, was implemented. All site personnel were blinded to the randomization sequence and the contents of the containers. The 10-mg dosage of JNJ-67953964 was chosen for use in this study because (1) a PET study demonstrated dose-dependent KOR occupancy and, at approximately peak occupancy (2.5 h after dosing), brain KORs were almost saturated with single doses of 10 mg or more (96.7% receptor occupancy) with a trough occupancy of >60%[24] and (2) 10 mg was the highest dosage where completed toxicity studies supported carrying out a trial as long as 8 weeks with daily dosing in humans.

Assessments. *POM and efficacy assessments.* fMRI during the Monetary Incentive Delay task. The primary outcome measure was the magnitude of fMRI-determined ventral striatal (including nucleus accumbens) activation during anticipation of rewards in the MID task[51]. This measure was chosen as a means of testing POM for KOR antagonism because it was used as a measure of function in brain reward circuitry found to be modulated by KOR antagonism in preclinical work[14,17–19] and it differentiated individuals with depression from healthy controls[30].

fMRI results for the MID task were obtained at baseline and at the end (week 8) of double-blind therapy, and the task was performed according to previous designs[51]. The task was administered in five task runs, each consisting of 24 trials. For each trial, participants were presented with one of three possible cues for 500 ms, followed by a fixation crosshair on a computer screen. These cues signaled whether the upcoming trial had the potential for monetary gain ($n = 40$; denoted +$), had the potential for monetary loss ($n = 40$; denoted −$) or there was no possibility for monetary gain or loss ($n = 40$; denoted 0$). Participants were instructed that, on incentive trials, they could either gain or avoid losing money by pressing a button when presented with a red square target. On no-incentive trials, participants were instructed to still press the button as soon as the target appeared. Trial types were pseudorandomly ordered within each run. The duration of fixation following presentation of the cue was jittered between 2,250 and 3,750 ms, and the target was displayed for a period of 150 ms; 2,400–3,900 ms after target offset, participants were notified of how much money they had gained or lost on that trial.

Before testing, participants engaged in a training and practice run in the scanner. Task difficulty (that is, maximum allowable reaction time for both gain and loss trials) was titrated on the basis of reaction times collected during the practice session. Separate gain and loss reaction time standards were established to achieve approximately 70% success in each incentivized trial type.

MRI scan acquisition. All scans were conducted on research-dedicated 3.0-Tesla MRI scanners running the latest software version using an advanced 32-channel RF headcoil. However, the manufacturer and model type of the MRI scanners varied across sites, with the scanners including three Siemens Trios, one Siemens Verio, one Siemens Skyra and one GE MR750. A magnetic-resonance-compatible video-projection system with vision correction lenses, high-quality headphones and a button box was used for fMRI task presentation and response recording. The MRI acquisition sequence consisted of 15 s of localizer followed by gradient-echo echo-planar fMRI scans (axial; TR/TE, 2,000/30 ms; flip angle, 70 degrees; field of view (FOV), 25.6 cm; matrix, 64 × 64; 32 axial slices; acceleration factor, 2; voxel size, 4 × 4 × 4 mm; 137 fMRI time points + 4 dummy scans at the beginning (total of 141 points/TRs); total scan time, 4 min 42 s (141 × 2 s)) for each of five fMRI runs.

fMRI data processing was carried out with FEAT (FMRI expert analysis tool) v6.00, part of FSL (FMRIB's software library; https://www.fmrib.ox.ac.uk/fsl). Registration to high-resolution structural and/or standard space images was carried out with FLIRT[64,65]. Registration from high-resolution structural to standard space was then further refined with FNIRT nonlinear registration[66]. We employed the MNI152 normalization template. The following prestatistics processing was applied: motion correction using MCFLIRT[65]; slice timing correction using Fourier space time-series phase shifting; non-brain removal using BET[67]; spatial smoothing using a Gaussian kernel of full width at half-maximum (FWHM) = 5 mm; grand mean intensity normalization of the entire four-dimensional dataset by a single multiplicative factor; and high-pass temporal filtering (Gaussian weighted least-squares straight line fitting, with sigma = 45.0 s). Time-series statistical analysis was carried out by using FILM with local autocorrelation correction[68]. Higher-level analysis was carried out with a fixed-effects model, by forcing the random-effects variance to zero in FLAME (FMRIB's local analysis of mixed effects)[69,70]. The primary contrast of interest was averaged activation during reward anticipation (time points from onset of reward type cue to onset of target cue), for the contrast cued reward > cued non-reward. The primary outcome measure was obtained for an a priori–specified bilateral non-thresholded ventral striatal area mask, defined by the Harvard–Oxford Subcortical Atlas, and involved an a priori contrast using GLM of averaged z statistics for all voxels within the ROI. As exploratory outcomes, we also computed the maximum activation during reward anticipation (time points from onset of reward type cue to onset of target cue) for the contrast cued reward > cued non-reward and the mean and maximum activation during loss anticipation (time points from onset of loss type cue to onset of target cue) for the contrast cued loss > cued non-reward, for the a priori–specified ventral striatal area mask.

We instituted a set of quality-control procedures to standardize fMRI methods across sites. First, all sites were provided with the same E-Prime files for running the MID fMRI protocol and written materials outlining detailed fMRI methods. We also presented the fMRI methods in detail several times to all sites. In addition, our fMRI leads (A.S. and M.S.) provided one-on-one consultation to site fMRI personnel. We also required all sites to obtain and upload agar phantom scans to our central fMRI data analysis site (the laboratory of A.S. at Duke) both to qualify to begin to enroll participants and regularly throughout the study. Specifications for the agar phantom scans and their review and analysis were carried out as in the FIRST-BIRN multisite fMRI study quality assurance protocol[71]. The phantom scan raw data were reviewed, and estimates of signal-to-noise ratio (SNR) and signal-to-fluctuation noise ratio (SFNR) were generated. The signal image was the voxel-by-voxel average across all of the images. The fluctuation noise image was the s.d. of the residuals resulting from detrending the time series across all images for each voxel using a second-order polynomial. The SNFR was then generated by computing the voxel-by-voxel ratio of the signal image and the temporal

fluctuation image. A summary value for the SNFR was also computed taking the average of the SNFR across a 21×21 voxel ROI in the center of the SNFR image. The SNR computation started with creating separate sums of the odd- and even-numbered images and taking the difference between these two sums. The SNR was the ratio of the average of this difference image taken over a 21×21 voxel ROI in the center of the image to the square root of the variance of this difference within the same 21×21 voxel region divided by 198 time points. To compute the percent fluctuation and drift, first the average across images within a 21×21 voxel ROI in the center of the image was computed, generating a time series of average intensity. Next, the time series was fit with a second-order polynomial. The percent fluctuation was computed as 100 times the s.d. of the residuals of this fit divided by the average intensity. The percent drift was generated by dividing the difference between the maximum and minimum fit values by the average signal intensity and multiplying by 100.

The submitted scans were reviewed for artifacts and fluctuations over time, and the statistics derived from these scans were assessed for the degree of deviation from the mean values. Artifacts, systematic drift or substantial deviation based on the judgment of the MRI team reviewer led to contact with the site and implementation of a plan for correcting the identified problems. Lastly, we made on-site help available to the sites, which was required for one of the sites to be able to successfully implement the procedures and meet our criteria for standardization.

Snaith–Hamilton Pleasure Scale. The SHAPS[44] was used to screen participants, was a secondary outcome measure and was obtained at every visit. It is a 14-item questionnaire used to assess anhedonia covering four domains of hedonic experience: interest/pastimes, social interaction, sensory experience and food/drink. It asks participants to agree or disagree with statements of hedonic response in pleasurable situations (for example, 'I would enjoy my favorite television or radio program') on the basis of their experience in the 'last few days'. Four responses are possible—strongly disagree, disagree, agree or strongly agree. A total score can be derived by summing the responses to each item. Items answered with 'strongly agree' were coded as 1, while a 'strongly disagree' response was assigned a score of 4. Therefore, scores on the SHAPS can range from 14 to 56, with higher scores corresponding to higher levels of anhedonia[45]. This scale has shown adequate overall psychometric properties, including convergent validity[44,46,47], discriminant validity[45] and test–retest reliability[45]. Another important consideration that supports the use of the SHAPS in this study is that it is the only anhedonia measure found to significantly improve with the administration of treatments in clinical trials[72,73].

Probabilistic Reward Task. The PRT was a secondary outcome measure obtained at baseline and after 8 weeks of double-blind treatment. The PRT was designed to objectively assess the propensity to modulate behavior as a function of reinforcement history and has been found to reflect reward-related function in multiple independent samples[32,33,35–38]. Participants completed two blocks of 100 trials where they determined whether a briefly presented mouth on a cartoon face was 'long' or 'short' and reported their decision by pressing one of two corresponding keys on a computer keyboard (z or /). Importantly, the brief presentation time (100 ms) and the minimal difference in length between the two target stimuli (11.5 versus 13 mm) made it difficult for participants to distinguish the stimuli. Moreover, an asymmetrical reinforcement ratio was implemented across the two blocks so that one of the two stimuli (the 'rich' stimulus) was consistently rewarded ('Correct!! You Won 20 Cents') three times more frequently than the 'lean' stimulus (30 versus 10 times per block). Reinforcement allocation and key assignments were counterbalanced across participants. Participants were instructed to respond as quickly and accurately as possible to maximize monetary rewards and that not all correct responses were followed by rewards. Owing to the asymmetrical reinforcement schedule, performance in the PRT can be decomposed into response bias (log b) and discriminability (log d), which were computed as:

$$\log b = \frac{1}{2}\log\left(\frac{\text{rich}^{*}_{\text{correct}}\text{lean}_{\text{incorrect}}}{\text{rich}^{*}_{\text{incorrect}}\text{lean}_{\text{correct}}}\right) \text{ and } \log d = \frac{1}{2}\log\left(\frac{\text{rich}^{*}_{\text{correct}}\text{lean}_{\text{correct}}}{\text{rich}^{*}_{\text{incorrect}}\text{lean}_{\text{incorrect}}}\right)$$

To allow calculations in cases with a zero in one cell of the formula, 0.5 was added to every cell of the detection matrix[34].

Diminished preference for the rich stimulus (a decrease in response bias) has been found in individuals with increased depressive symptoms[33] and current MDD[32,33,36], particularly in those with elevated anhedonic symptoms[32,33] or melancholic depression[39], as well as in youth reporting anhedonia across various DSM diagnoses[40]. Before response bias and discriminability scores were computed, quality-control checks were performed with blinding to drug randomization by using a priori–defined cutoffs applied in recent PRT studies (Supplementary Note)[38].

The change in response bias from block 1 to block 2 served as a secondary outcome measure, which was tested by evaluating the three-way interaction of treatment arm(JNJ, placebo)×block(1, 2)×time(baseline, treatment week 8). This measure captures the total amount of reward learning across blocks. As an exploratory outcome, we also evaluated the treatment arm(JNJ, placebo)×time(baseline, treatment week 8) effect, which is computed as part of the analysis of the treatment group×block×time interaction. This effect probed changes in overall response bias (that is, averaged across both blocks) as a function of treatment and might be especially sensitive for designs (such as in the current study) involving only two blocks of the PRT, which might not allow response bias to grow throughout the blocks as strongly as in previous studies that used three blocks[32,33,41].

Resting-state delta EEG current density in the rostral anterior cingulate. We obtained resting-state, eyes-closed quantitative 32-channel EEG (QEEG) to provide an additional, exploratory circuit-based measure of hedonic function. From these data, we computed current density by using low-resolution electromagnetic tomography analysis (LORETA) on the basis of evidence that, when using this method, greater resting EEG delta (1.5–6 Hz) current density (that is, lower brain activity) in the rostral anterior cingulate is correlated with higher anhedonia scores among healthy individuals[74]. The recording electrodes were placed according to the Modified International 10–20 System. Right infra-orbital and left outer canthus electrodes were used to monitor eye movements. Data were obtained with a referential montage employing a linked-ear reference by using a digitization rate of 256 Hz. Electrode impedances were maintained at below 5 kΩ during the recordings, and standard square-wave and biocalibrations were performed. Data collection occurred for 20 min (10 min with the eyes closed and 10 min with the eyes open) following calibration with filter settings of 0.5 and 70 Hz. Manual epoch-by-epoch artifacting was carried out by the central Duke QEEG Core with blinding to participant identity and study time point. QEEG analysis was only carried out if a minimum of 30 s of waking, artifact-free data were available.

Effort Expenditure for Rewards Task. The EEfRT[75], an exploratory behavioral measure for the study, was intended to assess the motivation to pursue rewards, one important dimension of reward-related function. It has been reported to be correlated with several anhedonia scales[75]. The EEfRT task is a multitrial game in which participants are given an opportunity on each trial to choose between two different task difficulty levels to obtain monetary rewards. For additional information about EEfRT methods, see the Supplementary Note.

Visual Analog Scale of Anhedonia. The VAS–Anhedonia is a standard VAS assessment of anhedonia severity that was included because it provides a global anhedonia indicator that takes very little time to obtain and that was found to be sensitive to change with treatment in a previous placebo-controlled trial in alcohol-dependent individuals[72]. The test consists of making a rating on a 100-mm scale in response to the directive 'Make a mark on the line below that indicates how much pleasure you experience from food, sexual behavior, and meeting friends'. At the left end of the scale is the anchor 'No Pleasure' and at the right end of the scale is the anchor 'Extreme Pleasure'.

Temporal Experience of Pleasure Scale. We also included the TEPS as an exploratory measure because it provides different information about reward-related function than the SHAPS and has been found to be correlated with activation in the key circuits of interest (the nucleus accumbens and putamen) in MID task-related fMRI[76]. The TEPS is an 18-item self-report measurement of anticipatory (10 items) and consummatory (8 items) components of anhedonia, consisting of a series of statements that must be rated according to how accurate they are for the individual[76].

The Hamilton Rating Scale for Depression. The HAM-D 17-item version was included in exploratory analysis to provide confirmatory support for changes in depression severity with treatment[48]. This interviewer-administered semistructured interview is one of the most widely used instruments in depression treatment studies.

The Hamilton Rating Scale for Anxiety. The HAM-A is a rating scale designed to measure the severity of anxiety symptoms[49]. It is widely used in both clinical and research settings. The scale consists of 14 items, each defined by a series of symptoms, and measures both psychic anxiety (mental agitation and psychological distress) and somatic anxiety (physical complaints related to anxiety). It has been demonstrated to have acceptable reliability, validity and sensitivity to change[49]. Each item is scored on a scale of 0 (not present) to 4 (severe), with a total score range of 0–56, where a score of 9–16 indicates mild severity, 18–24 indicates mild to moderate severity and 25–30 indicates moderate severity to severe anxiety; scores greater than 30 indicate severe anxiety. This instrument was included in secondary analysis to provide confirmatory support for changes in anxiety severity with treatment.

The Cognitive and Physical Functioning Questionnaire. The CPFQ is a seven-item self-report instrument included in exploratory analysis intended to be a brief scale for measuring cognitive and executive dysfunction in individuals with mood and anxiety disorders[77]. This scale has been demonstrated to have strong internal consistency, good temporal stability and sensitivity to change with treatment[77].

Clinical Global Impression. The CGI-S is a widely administered clinician-rated global measure of overall illness severity. Individuals are rated on a scale from 1 to 7, where 1 corresponds to 'Normal, Not at All Ill', 2 is 'Borderline Mentally Ill', the anchor for 3 is 'Mildly Ill', the anchor for 4 is 'Moderately Ill', 5 is 'Markedly Ill',

6 is 'Severely Ill' and 7 is 'Among the Most Extremely Ill Patients'. The CGI-I is a widely administered clinician-rated global measure of the degree of improvement from the initial assessment in overall illness severity. Individuals are rated on a scale from 1 to 7, where 1 corresponds to 'Very Much Improved', 2 is 'Much Improved', the anchor for 3 is 'Minimally Improved', the anchor for 4 is 'No Change', 5 is 'Minimally Worse', 6 is 'Much Worse' and 7 is 'Very Much Worse'. Both the CGI-S and CGI-I were administered at all visits. During the course of the trial, participants for whom the CGI-I was greater than 5 were removed from the study and appropriate care was given, for safety purposes.

Safety assessments. Safety assessments carried out at all visits included clinical evaluation of adverse events, adverse effect assessment with the Patient-Reported Inventory of Side-Effects (PRISE), vital signs (height, weight, blood pressure and pulse), urine drug screen, complete blood count with differential, electrolytes, comprehensive metabolic panel including liver function tests, thyroid function tests, urinalysis, CGI-I and ECG. Physical examination was carried out at baseline and at the end of double-blind treatment. A β-hCG serum pregnancy test was obtained during screening. Suicidality was assessed at every visit by using the CSSRS[63]. Tests for assessment of gastric adverse events were obtained at baseline and at weeks 4 and 8 of double-blind treatment and included measurement of gastrin and pepsinogen I and II levels. These gastric tests were obtained for the first half of the study on the basis of an FDA recommendation for monitoring, after which approval was obtained from the FDA, the NIMH DSMB and the relevant institutional review boards to drop this requirement.

Oversight. The trial was conducted in accordance with International Conference on Harmonisation Good Clinical Practice guidelines and was approved by the relevant institutional review boards. Written informed consent was provided by the patients or their legal representatives. Data were collected and analyzed by the investigators and interpreted by all authors. The first draft of the manuscript was prepared by the first author. All authors approved subsequent drafts and agreed to submit the manuscript for publication. The authors had full access to the trial data and vouch for the accuracy and completeness of the data and for the fidelity of the trial to the protocol. The trial was governed by a committee of site investigators and NIMH program officers.

Outcomes. *Primary outcome.* The primary outcome for evaluating POM that engaging the target (KOR antagonism) had the hypothesized effect on reward-related brain function was the baseline-corrected fMRI activation at the end of 8 weeks of double-blind treatment in an a priori bilateral non-thresholded ventral striatal area mask, defined by the Harvard–Oxford Subcortical Atlas, during anticipation of monetary gain in the MID task as contrasted with neutral (no-gain/no-loss) trials.

Owing to the experimental medicine approach taken, the driving factor leading to choice of the primary outcome measure was the neurobiological target. Accordingly, on the basis of the compelling preclinical literature available to us when we designed the study indicating that KOR antagonism releases inhibition on DA neurons and increases nucleus accumbens function[14–21], we chose reward-related activation in the nucleus accumbens as our primary target. Having committed to this mechanism and target, we sought an fMRI paradigm that would reliably engage the nucleus accumbens in response to rewards. On the basis of the early work of Knutson[50,52–54] and others[55,56], use of the MID task in anticipation of gains was selected. We note that, although exceptions exist[51,57], previous studies in both MDD and psychiatrically healthy samples available to us when the study was designed have linked reward-related ventral striatal activation and anhedonic symptoms[74], and this relationship has been replicated in more recent studies[58,59]. Further, recent meta-analyses of fMRI data collected among healthy control individuals have confirmed that the ventral striatum (nucleus accumbens) is reliably recruited during reward anticipation[60,78].

This choice was further supported by the fact that there were data from a previous treatment study (open-label escitalopram administered to 15 individuals with MDD and 15 controls) where outcome was assessed with fMRI-determined ventral striatal activation during anticipation of gains and losses in the MID task[30]. That study provided a set of findings that differed for ventral striatal activation in anticipation of gains and losses but suggested overall the promise of both measures. Findings suggesting the relative utility of ventral striatal activation in anticipation of loss include that the baseline Beck Depression Inventory (BDI) anhedonia item score was significantly related to ventral striatal loss anticipation and not gain and that escitalopram led to a significant change in ventral striatal loss-anticipation-related activity but not activity in anticipation of gains. However, a number of findings supported our choice to employ ventral striatal activation in anticipation of gains in our study. These included the following: (1) participants with depression had significantly smaller ventral striatal activation in anticipation of gain than individuals in the control group before treatment with escitalopram ($P=0.025$) but not after treatment; (2) there was a trend for a group × time interaction with treatment with ventral striatal activation in anticipation of gain ($P<0.07$) in this study with only 15 individuals per group; and (3) there was a trend toward decreased BDI score over treatment being correlated with increases in ventral striatal activation in anticipation of gain ($r=0.54$, $P=0.058$).

Secondary outcomes. A priori–specified secondary outcomes included:

1. A clinical measure of anhedonia, the SHAPS;
2. A behavioral measure of reward-related function, the change in response bias from block 1 to block 2 in the PRT (evaluated by testing the treatment arm × block × time interaction).

Exploratory outcomes. All other outcomes were considered in exploratory analyses, including the PRT group × time interaction, resting-state delta EEG current density in the rostral anterior cingulate, TEPS, EEfRT, VAS-Anhedonia, HAM-D, HAM-A, CPFQ and CGI.

Statistical analysis. *Sample size.* We powered this study to detect an effect size of 0.5. Our capacity to estimate the expected effect size on our primary outcome measure was limited because there has been only one treatment study carried out with this measure so far[30]. This previous study, referred to above, involved open-label escitalopram treatment of 15 individuals with MDD and 15 control individuals. Pre- to post-treatment effects and the group × time interaction suggested an effect size of 0.88 or higher with our primary outcome measure. It was understood that estimates of power based on the previous treatment study were likely overly optimistic because it was an open-label treatment study and only included 15 individuals with depression. As a result, we assumed a lower effect size of 0.50. On this basis, we planned to enroll 90 total participants, with the conservative expectation that we could have incomplete data on up to 20% of participants owing to dropout or loss of data due to factors such as poor scan quality. This would make available at least 72 individuals for analyses, which provides 80% power to detect an effect size of 0.5 at the significance level of alpha = 0.05 in a one-tailed test of significance.

Efficacy analyses. We planned to carry out efficacy analyses in the ITT population and to perform one-sided statistical tests using a *P*-value threshold of 5% for statistical significance. Analyses consisted of mixed-effects models including baseline values, age, study site and sex as covariates. The mixed-effects models used a random intercept and fixed slopes model with compound symmetry structure and employed a maximum-likelihood estimation approach.

Site was included as an independent variable with analysis carried out employing centering based on Kraemer and Blasey[79] to account for variability among the study sites. Because of the number of covariates and sites relative to the sample size, examination of site × time and site × treatment arm × time interactions were not preplanned analyses.

The rationale for including baseline as a covariate in mixed-effects models is included in the Supplementary Note.

For each variable, the baseline-adjusted mean was computed for both treatment groups. We also computed JNJ-67953964 versus placebo group effect sizes (Hedges' *g*), which were calculated as (ME – MC)/s.d. pooled, where ME represents the adjusted mean of experimental treatment, MC represents the adjusted mean of the comparison treatment and s.d. pooled represents pooling of the standard deviations from within both groups. Hedges' *g* is similar to Cohen's *d* except that it employs a sample-size-weighted pooled s.d. whereas Cohen's *d* employs a pooled s.d.[28,31]. As a result, Hedges' *g* is believed to be a less biased measure of effect size with groups of unequal size and in datasets of limited size, which is why it was chosen for use in this study[28].

Multiple imputations were used in this study to account for missing data to verify the results for the primary endpoint for the ITT and as-treated populations. The per-protocol analyses did not employ multiple imputations. We performed multiple imputations of missing data by using the SAS PROC MI procedure employing the Markov chain Monte Carlo method with a single chain carrying out 20 imputations with a seed of 788, using the same covariates as for our mixed-effects model analysis. We then analyzed the complete datasets by using the MIXED procedure, and we analyzed the output from the two previous steps by using the MIANALYZE procedure. The hypothesized missing mechanism was 'missing at random'. We employed a missing-at-random model despite the fact that the reasons for discontinuation were known because the reasons for discontinuation did not suggest additional variables that could be included in the multiple-imputation models that would be likely to be predictive of missingness/discontinuation[80]. The reasons for discontinuation, outlined in the Supplementary Note, reflect a relatively even distribution across a number of different discontinuation reasons.

In addition to the above analyses, we carried out a set of post hoc exploratory analyses to follow up on the findings of the planned analyses. To evaluate the capacity to determine the degree with which response to treatment can be predicted from the baseline value of our primary outcome, we carried out an exploratory correlation analysis. Moreover, to determine the extent to which response to treatment can be predicted at the individual level, we carried out a logistic regression analysis determining the extent to which baseline ventral striatal activation predicted which individuals would be responders in terms of ventral striatal activation while controlling for treatment arm. Response was defined on the basis of a median split for the change in ventral striatal activation with treatment.

We also carried out exploratory post hoc correlation analyses controlling for treatment arm to determine the degree of relatedness of our primary outcome measure with the key secondary outcome measures, the SHAPS score and the PRT change from block 1 to block 2 in response bias.

Lastly, we carried out an exploratory post hoc analysis of the magnitude of the change in mean ventral striatal activation in anticipation of gain in response to treatment separately in the JNJ-67953964 and placebo groups. This was performed to better delineate the nature of the between-group effects found with this measure. This analysis consisted of determining the effect sizes (Hedges' *g*) for the change in our primary outcome measure for the JNJ-67953964 and placebo groups. We evaluated effect sizes rather than assessing statistical significance because our study was not powered to demonstrate statistical significance in such analyses, each of which included approximately half of the total sample size. Furthermore, unlike statistical significance, effect size is not dependent on sample size. An important consideration with respect to this analysis was that we assumed a priori that ventral striatal activation in anticipation of gain might tend to decrease over time. As a result, our plan was to compare the drug group with the placebo group and not carry out within-group statistical significance testing, because we were aware of the possibility that repeating the test after 8 weeks in the population studied might be accompanied by an adaptation effect that would be manifested in a tendency toward a diminished response in the second test session in both groups. The comparison to placebo was intended to mitigate this contingency. The reasons we believed there was a possibility that ventral striatal activation might decrease in the second test session versus the first are outlined in the Discussion.

No interim analyses were planned or carried out.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Study data have been posted to the NIH/NIMH data archive and are accessible by emailing NDAHelp@mail.nih.gov.

References

61. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders* 4th edn (American Psychiatric Association, 2009).
62. Sheehan, D. V. et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J. Clin. Psychiatry* **59**, 22–33 (1998).
63. Posner, K. Columbia–Suicide Severity Rating Scale (C-SSRS) http://www.cssrs.columbia.edu (2011).
64. Jenkinson, M., Bannister, P., Brady, M. & Smith, S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* **17**, 825–841 (2002).
65. Jenkinson, M. & Smith, S. A global optimisation method for robust affine registration of brain images. *Med. Image Anal.* **5**, 143–156 (2001).
66. Andersson, J. L. R., Jenkinson, M. & Smith, S. Non-linear registration, aka Spatial normalisation. FMRIB technical report TR07JA2 FMRIB Centre, 2007.
67. Smith, S. M. Fast robust automated brain extraction. *Hum. Brain Mapp.* **17**, 143–155 (2002).
68. Woolrich, M. W., Ripley, B. D., Brady, M. & Smith, S. M. Temporal autocorrelation in univariate linear modeling of fMRI data. *Neuroimage* **14**, 1370–1386 (2001).
69. Woolrich, M. W., Behrens, T. E. J., Beckmann, C. F., Jenkinson, M. & Smith, S. M. Multilevel linear modelling for fMRI group analysis using Bayesian inference. *Neuroimage* **21**, 1732–1747 (2004).
70. Woolrich, M. W. et al. Bayesian analysis of neuroimaging data in FSL. *Neuroimage* **45**, S173–S186 (2009).
71. Friedman, L. & Glover, G. H. Report on a multicenter fMRI quality assurance protocol. *J. Magn. Reson. Imaging* **23**, 827–839 (2006).
72. Di Giannantonio, M. & Martinotti, G. Anhedonia and major depression: the role of agomelatine. *Eur. Neuropsychopharmacol.* **22**, S505–S510 (2012).
73. Martinotti, G. et al. Acetyl-L-carnitine in the treatment of anhedonia, melancholic and negative symptoms in alcohol dependent subjects. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **35**, 953–958 (2011).
74. Wacker, J., Dillon, D. G. & Pizzagalli, D. A. The role of the nucleus accumbens and rostral anterior cingulate cortex in anhedonia: integration of resting EEG, fMRI, and volumetric techniques. *Neuroimage* **46**, 327–337 (2009).
75. Treadway, M. T., Buckholtz, J. W., Schwartzman, A. N., Lambert, W. E. & Zald, D. H. Worth the 'EEfRT'? The effort expenditure for rewards task as an objective measure of motivation and anhedonia. *PLoS One* **12**, e6598 (2009).
76. Gard, D. E., Gard-Germans, M., Kring, A. M. & Oliver, J. P. Anticipatory and consummatory components of the experience of pleasure: a scale development study. *J. Res. Personality* **40**, 1086–1102 (2006).
77. Fava, M., Iosifescu, D. V., Pedrelli, P. & Baer, L. Reliability and validity of the Massachusetts General Hospital cognitive and physical functioning questionnaire. *Psychother. Psychosom.* **78**, 91–97 (2009).
78. Wilson, R. P. et al. The neural substrate of reward anticipation in health: a meta-analysis of fMRI findings in the monetary incentive delay task. *Neuropsychol. Rev.* **28**, 496–506 (2018).
79. Kraemer, H. C. & Blasey, C. M. Centering in regression analyses: a strategy to prevent errors in statistical inference. *Int. J. Methods Psychiatr. Res.* **13**, 141–151 (2004).
80. Azur, M. J., Stuart, E. A., Frangakis, C. & Leaf, P. J. Multiple imputation by chained equations: what is it and how does it work? *Int. J. Methods Psychiatr. Res.* **20**, 40–49 (2011).

Acknowledgements

This project was supported by contract HHS-N271-2012-000006-I from the National Institute of Mental Health awarded to A.D.K. D.A.P. was partially supported by R37 MH068376. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the US government.

Author contributions

A.D.K. designed the study, provided oversight for all aspects of the study, analyzed the data and wrote the manuscript. W.Z.P. provided critical intellectual input for designing the study and contributed to writing the manuscript. D.A.P. provided critical intellectual input for designing the study, provided training, oversight and data analysis for the aspects of the study related to the behavioral tests, and co-wrote the manuscript. A.E.W. provided training, oversight and data analysis for the aspects of the study related to the behavioral tests. S.J.M., S.H.L., J.R.C., G.S. and W.G. contributed to designing the study, provided oversight for a study site where they led efforts in training and managing study personnel, recruiting participants and completing all protocol-specified study procedures including data collection with the participants, and contributed to writing the manuscript. R.D.W. and J.N. provided oversight for a study site where they led efforts in training and managing study personnel, recruiting participants and completing all protocol-specified study procedures including data collection with the participants and contributed to writing the manuscript. A.S. and M.S. provided oversight for collection and analysis of the MRI data and contributed to writing the manuscript. H.Y. served as statistician for the study, contributing to study design, carrying out data analysis and contributing to writing the manuscript. R.S.E.K. contributed to designing the study, provided data quality oversight and contributed to writing the manuscript. D.I., S.T.S., G.H. and K.G. participated in recruiting participants, completing all protocol-specified study procedures including data collection with the participants, and contributed to writing the manuscript.

Competing interests

A.D.K. has been a consultant for Adare, Eisai, Ferring, Galderma, Idorsia, Jazz, Janssen, Takeda, Merck, Neurocrine, Pernix, Physician's Seal, Evecxia and Sage Research and received support from the NIH, the Ray and Dagmar Dolby Family Fund, Janssen, Jazz. Axsome and Reveal Biosensors. D.A.P. has consulted for Akili Interactive Labs, BlackThorn Therapeutics, Boehringer Ingelheim, Compass Pathway, Otsuka Pharmaceuticals, Posit Science and Takeda Pharmaceuticals and received honoraria from Alkermes. D.A.P. has a financial interest in BlackThorn Therapeutics, which has licensed the copyright to the PRT through Harvard University. The interests of D.A.P. were reviewed and are managed by McLean Hospital and Partners HealthCare in accordance with their conflict-of-interest policies. No funding from these entities was used to support the current work, and all views expressed are solely those of the authors. S.J.M. has been a consultant for Allergan, Alkermes, Greenwich Biosciences, Clexio Biosciences, Intra-Cellular Therapies, Janssen, Perception Neuroscience, Sage Therapeutics, Signant Health and Seelos Therapeutics and received research support from Biohaven, Janssen, the NIH, NeuroRx and VistaGen Therapeutics, drug from Biohaven for NIMH-funded study and support from the Michael E. Debakey VA Medical Center (Houston, TX) for use of resources and facilities. G.S. has consulted for Allergan, Alkermes, AstraZeneca, Avanier Pharmaceuticals, Axsome Therapeutics, Biohaven Pharmaceuticals, Boehringer Ingelheim, Bristol-Myers Squibb, Clexio Biosciences, EMA-Wellness, Hoffman La-Roche, Intra-Cellular Therapies, Janssen, Merck, Naurex, Navitor Pharmaceuticals, Novartis, Noven Pharmaceuticals, Otsuka, Praxis Therapeutics, Sage Pharmaceuticals, Servier Pharmaceuticals, Taisho Pharmaceuticals, Teva, Valeant and Vistagen Therapeutics and received research funding from AstraZeneca, Bristol-Myers Squibb, Eli Lilly, Johnson & Johnson, Hoffman La-Roche, Merck, Naurex and Servier. G.S. has equity interests in Biohaven Pharmaceuticals and receives patent royalties from Biohaven. J.W.M. has consulted for Boehringer Ingelheim, Sage Therapeutics, Novartis, Allergan, Fortress Biotech, Janssen Research and Development, Medavante-Prophase and Global Medical Education (GME) and received research support from Avanir Pharmaceuticals. R.S.E.K. has consulted for or been a speaker or advisory board member for Abbvie, Acadia, Aeglea, Akebia, Akili, Alkermes, Allergan, ArmaGen, Astellas, Avanir, Axovant, Biogen, Boehringer Ingelheim, Cerecor, CoMentis, Critical Path Institute, FORUM, Gammon Howard & Zeszotarski, Global Medical Education (GME), GW Pharmaceuticals, Intracellular Therapeutics, Janssen, Kempharm, Lundbeck, Lysogene, MedScape, Mentis Cura, Merck, Merrakris Therapetics, Minerva Neurosciences, Mitsubishi, Montana State University, Monteris, Moscow Research Institute of Psychiatry, Neuralstem, Neuronix, Novartis, the New York State Office of Mental Health, Orygen, Otsuka, Paradigm Testing, Percept Solutions, Pfizer, Pharm-Olam, Regenix Bio, Reviva, Roche, Sangamo, Sanofi, SOBI, Six Degrees Medical,

Sunovion, Takeda, Targacept, Teague Rotenstreich Stanaland Fox & Holt, Thrombosis Research Institute, the University of Moscow, the University of Southern California, the University of Texas Southwestern Medical Center, WebMD and Wilson Therapeutics. R.S.E.K. has received research funding from the National Institute of Mental Health and Boehringer Ingelheim and royalties from Versions of the BAC testing battery, the MATRICS Battery (BACS Symbol Coding) and the Virtual Reality Functional Capacity Assessment Tool (VRFCAT). R.S.E.K. is a shareholder in VeraSci and Sengenix. S.H.L. contributed to this article while at Duke University, before joining the National Institute of Mental Health. The views expressed are her own and do not necessarily represent the views of the National Institutes of Health or the US government. S.H.L. is a co-inventor on a patent for TMS Technology, unrelated to this manuscript. J.N. has received research funding from Janssen and Assurex. A.S. has received research support from the NIH and GE Healthcare. W.G. has consulted for Biohaven Pharmaceuticals and received research funding from the NIH, the Simons Foundation and Biohaven Pharmaceuticals Device Donation: Medtronic. S.T.S. has consulted for Otsuka Pharmaceuticals, Neurocrine Biosciences, Jazz, Teva and the Centers of Psychiatric Excellence and Continuous Precision Medicine and received research support from Otsuka Pharmaceuticals and Veterans' Affairs grant IK2CX001397. K.G. has been on the advisory board of Sunovion and Otsuka Pharmaceuticals and received research funding from AstraZeneca, the Cleveland Foundation and the Brain and Behavioral Research Foundation. K.G. has been a speaker for AstraZeneca, Pfizer and Sunovion. W.Z.P. has been on the advisory board and/or consulted for Takeda, Lilly, Praxis, Astellas, Otsuka and Noven. W.Z.P. has been on the DSMB for Agene-Bio and has stock ownership in Merck.

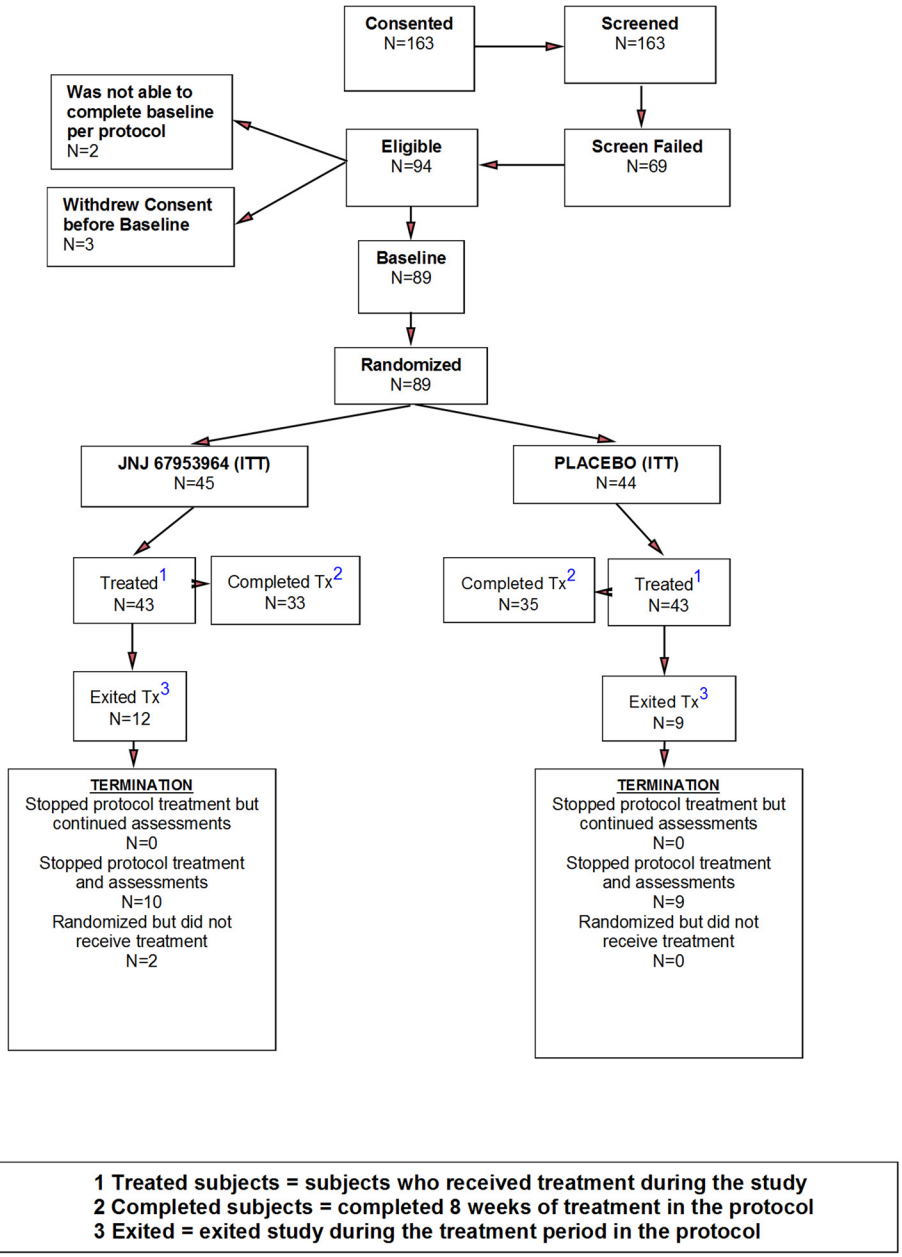
Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41591-020-0806-7.

Supplementary information is available for this paper at https://doi.org/10.1038/s41591-020-0806-7.

Correspondence and requests for materials should be addressed to A.D.K.

Peer review information Kate Gao was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.



Extended Data Fig. 1 | Consort diagram. Study consort diagram.

nature research

Corresponding author(s): Andrew D. Krystal, MD, MS

Last updated by author(s): 1/29/2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	InForm,™ a thin client browser based internet-accessed commercially available Electronic Data Capture (EDC) system that uses HTML only on the client was used for site data entry.
Data analysis	Data Analysis was carried out with SAS 9.4

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The datasets generated and analyzed during the current study are available in the National Institute of Mental Health (NIMH) Data Archive repository and can be accessed by contacting the NIMH Data Archive at the following email address: NDAHelp@mail.nih.gov

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	A power analysis was carried out to determine sample size.
Data exclusions	No data were excluded from the analyses
Replication	No attempts at replication were carried out.
Randomization	Random assignment to groups was carried out for all subjects.
Blinding	All investigators were blinded to group membership during data collection and analysis.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☐	☒ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☐	☒ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	The study cohort included 89 individuals who had clinically significant anhedonia as defined by a Snaith Hamilton Pleasure Scale (SHAPS) score of at least 20 and currently met DSM-IV TR diagnostic criteria for Major Depressive Disorder (MDD), Bipolar I or II Depressed, Generalized Anxiety Disorder, Social Phobia, Panic Disorder, or Post Traumatic Stress Disorder based on the Mini-International Neuropsychiatric interview. Potential subjects were excluded if they: were expected to require any hospitalization during the course of the study; had a history of a psychotic disorder, current manic or mixed episode, autism spectrum disorders, mental retardation; met DSM-IV-TR criteria for substance abuse within the last 3 months or substance dependence within the last 6 months; had a history of unstable or untreated serious medical condition; had active suicidal intent or plan, or history of attempt within the past 3 months based on physician evaluation and the Columbia Suicide Severity Rating Scale (C-SSRS);43 used any medication with significant central nervous system effects including antidepressants, antipsychotics, anxiolytics, anticonvulsants, mood stabilizing agents, muscle relaxants, centrally acting anti-histaminergics, stimulants or insomnia medications within 5 half-lives of baseline or at any time during the study; used any medication that is primarily metabolized by Cytochrome P450 2C8 within 14 days of baseline or at any time during the study; had any contraindications to the MRI procedures; had a positive urine drug screen at any time during the study; used any investigational medication within 3 months; had a history of gastric disease (including peptic ulcer disease, gastritis, upper GI bleeding, or any GI precancerous condition), had current clinically evident GI complaints; had a positive urea breath test (exclusionary in the first half of the study, after which approval was obtained from FDA, the NIMH DSMB, and the relevant Institutional Review Boards to drop this requirement); were less than 21 or more than 65 years of age; current use of a proton pump inhibitor or histamine 2 blocker, or a history of chronic non-steroidal anti-inflammatory drugs use; history of use of Salvia divinorum or use of Salvia divinorum at any time during the study; any smoking of cigarettes or use of other nicotine containing products within the last month or at any time during the study; or were pregnant or lactating. The 89 individuals in the study cohort had an average age of 39.5 years (S.D.=13.2), were 62.9% female, were 67.8% caucasian, 20.7% african american, were 3.4% asian, were 1.1% American Indian/Alaskan Native, were 11.6% hispanic in origin, had a mean BMI of 28.7 (S.D.=6.2), had a mean weight of 180.6 lbs (S.D.=41.9), had a mean SHAPS score of 34.9 (S.D.=7.4), had a mean Hamilton Depression Rating Scale score of 15.6 (S.D.=5.6) and a mean Hamilton Anxiety Rating Scale Score of 15.5 (S.D.=6.2).

Recruitment	Subjects were recruited via various methods across the 6 centers participating in this study. Methods including posting flyers, internet, radio, and newspaper advertisements, subject databases, and electronic medical record searches. There was no known self-selection bias.
Ethics oversight	Institutional Review Boards of Duke University, Yale University, Baylor University, Mt. Sinai University, Indiana University, Case-Western Reserve University. Study oversight was also provided by the NIMH Data Safety Monitoring Board

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	ClinicalTrials.gov Identifier: NCT02218736
Study protocol	Full Trial Protocol was submitted with manuscript
Data collection	The trial was conducted at six centers in the United States (Duke University, Yale University, Icahn School of Medicine at Mount Sinai, Baylor College of Medicine, Indiana University, and Case Western Reserve University) from September 2015 through October 2017
Outcomes	The apriori specified Primary Outcome Measure was task-related fMRI ventral striatal (e.g., nucleus accumbens) activation occurring with reward anticipation during the Monetary Incentive Delay (MID) Task. Two secondary outcome measures were apriori specified consisting of the total score on the Snaith-Hamilton Pleasure Scale (SHAPS) and the change in response Response Bias from Block 1 to Block 2 on the Probabilistic Reward Task (PRT). Exploratory outcome measures consisted of ventral striatal fMRI activation during anticipation of loss during the MID Task, resting state delta EEG current density in the rostral anterior cingulate, resting state fMRI connectivity, Effort-Expenditure for Rewards Task (EEfRT) score, Visual Analogue Scale for Anhedonia (VAS), The Temporal Experience of Pleasure Scale (TEPS) total score and consummatory and anticipatory subscores, the Hamilton Depression Rating Scale (HAM-D) score, the Hamilton Anxiety Scale (HAM-A) score, the Cognitive and Physical Functioning Questionnaire (CPFQ) score, Clinical Global Impression – Severity (CGI-S), and Clinical Global Impression – Improvement (CGI-I).

Magnetic resonance imaging

Experimental design

Design type	Event-Related
Design specifications	The Monetary Incentive Delay Test fMRI was administered in five task runs each consisting of 24 trials.
Behavioral performance measures	During the task, subjects pressed a button when presented with a red square target. Reaction time data was collected and means and standard deviations were reviewed across subjects to establish that subjects were performing the task as expected.

Acquisition

Imaging type(s)	Functional
Field strength	3.0 Tesla
Sequence & imaging parameters	The MRI acquisition sequence consisted of 15 sec of Localizer followed by Gradient-echo echo-planer fMRI scans, axial, TR/TE: 2000/30 ms, flip angle: 70 deg, FOV: 25.6 cm, matrix: 64x64, 32 axial slices, acceleration factor = 2, voxel size: 4x4x4 mm, 137 fMRI time points + 4 dummy scans at the beginning (total 141 points/TRs), total scan time 4 min 42 seconds (141 x 2 s) for each of five fMRI runs.
Area of acquisition	Whole brain scanning was carried out. The primary outcome measure was obtained for an a priori specified bilateral non-thresholded ventral striatal area mask, defined by the Harvard-Oxford Subcortical Atlas

Diffusion MRI ☐ Used ☒ Not used

Preprocessing

Preprocessing software	FSL version 5.0.6, FEAT version 6.0
Normalization	FMRI data processing was carried out using FEAT (FMRI Expert Analysis Tool) Version 6.00, part of FSL (FMRIB's Software Library, www.fmrib.ox.ac.uk/fsl). Registration to high resolution structural and/or standard space images was carried out using FLIRT [Jenkinson 2001, 2002]. Registration from high resolution structural to standard space was then further refined using FNIRT nonlinear registration [Andersson 2007a, 2007b].
Normalization template	MNI152

Noise and artifact removal

FMRI data processing was carried out using FEAT (FMRI Expert Analysis Tool) Version 6.00, part of FSL (FMRIB's Software Library, www.fmrib.ox.ac.uk/fsl). The following pre-statistics processing was applied; motion correction using MCFLIRT [Jenkinson 2002]; slice-timing correction using Fourier-space time-series phase-shifting; non-brain removal using BET [Smith 2002]; spatial smoothing using a Gaussian kernel of FWHM 5mm; grand-mean intensity normalisation of the entire 4D dataset by a single multiplicative factor; highpass temporal filtering (Gaussian-weighted least-squares straight line fitting, with sigma=45.0s)

Volume censoring

No volume censoring was done.

Statistical modeling & inference

Model type and settings

1st Level individual :
FMRI data processing was carried out using FEAT (FMRI Expert Analysis Tool) Version 6.00, part of FSL (FMRIB's Software Library, www.fmrib.ox.ac.uk/fsl). Time-series statistical analysis was carried out using FILM with local autocorrelation correction [Woolrich 2001].

2nd Level individual :
FMRI data processing was carried out using FEAT (FMRI Expert Analysis Tool) Version 6.00, part of FSL (FMRIB's Software Library, www.fmrib.ox.ac.uk/fsl). Higher-level analysis was carried out using a fixed effects model, by forcing the random effects variance to zero in FLAME (FMRIB's Local Analysis of Mixed Effects) [Beckmann 2003, Woolrich 2004, Woolrich 2008].

Effect(s) tested

The primary contrast of interest was averaged activation during reward anticipation (time points from onset of reward type cue to onset of target cue), for the contrast cued reward > cued non-reward.

Specify type of analysis: ☐ Whole brain ☒ ROI-based ☐ Both

Anatomical location(s)

The primary outcome measure was obtained for an a priori specified bilateral non-thresholded ventral striatal area mask, defined by the Harvard-Oxford Subcortical Atlas

Statistic type for inference
(See Eklund et al. 2016)

A priori contrast using GLM of averaged z-statistic of all voxels within the ROI.

Correction

No correction performed for a priori ROI contrast.

Models & analysis

n/a	Involved in the study
☒	☐ Functional and/or effective connectivity
☒	☐ Graph analysis
☒	☐ Multivariate modeling or predictive analysis

nature medicine

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41591-020-0806-7

In the format provided by the authors and unedited.

A randomized proof-of-mechanism trial applying the 'fast-fail' approach to evaluating κ-opioid antagonism as a treatment for anhedonia

Andrew D. Krystal[1,2]✉, Diego A. Pizzagalli[3], Moria Smoski[2], Sanjay J. Mathew[4,5], John Nurnberger Jr[6], Sarah H. Lisanby[7], Dan Iosifescu[8], James W. Murrough[9], Hongqiu Yang[10], Richard D. Weiner[2], Joseph R. Calabrese[11], Gerard Sanacora[12], Gretchen Hermes[12], Richard S. E. Keefe[2], Allen Song[2], Wayne Goodman[4], Steven T. Szabo[2,13], Alexis E. Whitton[3,14], Keming Gao[11] and William Z. Potter[7]

[1]Department of Psychiatry, University of California, San Francisco, San Francisco, CA, USA. [2]Departments of Psychiatry and Behavioral Sciences and Radiology, Duke University School of Medicine, Durham, NC, USA. [3]McLean Hospital, Harvard Medical School, Belmont, MA, USA. [4]Department of Psychiatry and Behavioral Sciences, Baylor College of Medicine, Houston, TX, USA. [5]Michael E. Debakey VA Medical Center, Houston, TX, USA. [6]Departments of Psychiatry and Neurobiology, Indiana University School of Medicine, Indianapolis, IN, USA. [7]National Institute of Mental Health, Bethesda, MD, USA. [8]Department of Psychiatry, New York University School of Medicine, New York, NY, USA. [9]Department of Psychiatry and Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [10]Duke Clinical Research Institute, Durham, NC, USA. [11]Department of Psychiatry, Case Western Reserve School of Medicine, Cleveland, OH, USA. [12]Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA. [13]Durham VA Health Care System, Durham, NC, USA. [14]The University of Sydney, Sydney, New South Wales, Australia. ✉e-mail: andrew.krystal@ucsf.edu

SUPPLEMENTARY INFORMATION:

Reasons for Dropouts

As described in the main text, the completer population consisted of 33 subjects in the JNJ-67953964 group and 35 placebo subjects. The reasons for dropout during double-blind treatment in the JNJ-67953964 group were: increase in depression symptoms (N=3); Hurricane Harvey prevented coming to the site and receiving study drug or led subject to have to leave town (N=2); subject unable to schedule visit within allowed time window (N=1); increase in anxiety symptoms (N=1); subject was lost to follow-up despite multiple attempts to contact (N=1); subject became pregnant (N=1); and subject had to start excluded medication for medical management of worsening of problem which predated study participation (N=1). The reasons for dropout in the placebo group were: increase in depression symptoms (N=3); subject was lost to follow-up despite multiple attempts to contact (N=2); increase in anxiety symptoms (N=1); subject dropped out due to developing back pain (N=1); and subject developed worsening of seasonal allergy symptoms and required an excluded medication (N=1).

A priori defined Quality Control Cutoff for the Probabilistic Reward Task

Quality control (QC) evaluations were performed blindly to Treatment Arm assignment and automatically using predefined QC cutoffs. Specifically, participants were excluded if any of the following QC were met:

1) Less than 80 valid trials in each block (i.e., more than 20% outlier responses). Outlier responses were defined in two steps:
    a. RT shorter than 150 ms or greater than 2,500 ms; and
    b. log-transformed RT exceeding the participant's mean ± 3SD.
2) less than 20 rich rewards or less than 6 lean rewards in each block;
3) rich-to-lean reward ratio < 2.0 in any block;

Additional Effort Expenditure for Rewards Task (EEfRT) Methods:

For all trials in the EEfRT, participants make repeated manual button presses within a short period of time. Each button press raises the level of a virtual “bar” viewed onscreen by the participant. Participants are eligible to win the money allotted for each trial if they raise the bar to the “top” within the prescribed time period. Each trial presents the subject with a choice between two levels of task difficulty, a ‘hard task’ and an ‘easy task.’ Successful completion of hard-task trials requires the subject to make 100 button presses, using the non-dominant little finger within 21 seconds, while successful completion of easy-task trials requires the subject to make 30 button presses, using the dominant index finger within 7 seconds. For easy-task trials, subjects are eligible to win the same amount, $1.00, on each trial if they successfully complete the task. For hard-task choices, subjects are eligible to win higher amounts that vary per trial within a range of $1.24 – $4.30 (“reward magnitude”). Subjects are not guaranteed to win the reward if they complete the task; some trials are “win” trials, in which the subject receive the stated reward amount, while others are “no win” trials, in which the subject receives no money for that trial. To help subjects determine which trials are more likely to be win trials, subjects are provided with accurate probability cues at the beginning of each trial. Trials have three levels of probability: “high” 88% probability of being a win trial, “medium” 50% and “low” 12%. Probability levels always apply to both the hard task and easy task, and there are equal proportions of each probability level across the experiment. Each

level of probability appears once in conjunction with each level of reward value for the hard task. All subjects receive trials presented in randomized order.

Rationale for Including Baseline As a Covariate In Mixed Effects Models:

Baseline was included as a covariate in mixed effects models to address the problem that differences between groups in baseline values of the outcome measure can negatively affect the trajectories of different treatment arms. This issue is critical when analyzing longitudinal data for two or more distinct groups with mixed effects models.[1,2] Including baseline as a covariate allows a comparison of the trajectories in the groups with the same baseline value for the outcome measure.[2] This is not achieved by the random intercept in the mixed effects models, which captures variations in overall tendencies that are not informed by known, measured differences between subjects at baseline. Further, achieving the goal of comparing trajectories between groups related to the same baseline value of the outcome conforms to the recommendations of the European Medicines Agency (EMA) who state in their “Guideline on Adjustment for Baseline Covariates in Clinical Trials” that when there is an association between baseline values and the outcome, adjustment for that difference generally improves the efficiency of the analysis and avoids conditional bias.[3] The approach taken of including baseline as a covariate is also in keeping with the specific recommendation of the EMA: “If a baseline value of a continuous primary outcome measure is available, then this should usually be included as a covariate.”.[3]

Results of Analysis Carried Out Without Controlling for Baseline Values:

Analyses were repeated for the primary and key secondary outcomes where statistically significant effects of treatment were found (SHAPS). A statistically significant treatment (JNJ-67953964 vs placebo) by time effect was found for the primary outcome measure (mean ventral striatal activation in anticipation of gain) when analysis when mixed-effects model analysis was carried out without controlling for baseline mean ventral striatal activation in anticipation of gain centered about its mean as a covariate ($F=1.9$; $p<0.027$). A statistically significant effect was not found when the mixed-effects model analysis was repeated for the SHAPS without controlling for baseline SHAPS score centered about its mean ($F=0.48$; $p=0.31$).

Consideration of the Relative Size of the VAS Anhedonia Scale JNJ-67953964 vs. Placebo Effect:

The VAS Anhedonia scale was among the exploratory clinical measures included in this study. While there was a tendency for greater improvement with JNJ-67953964 than placebo on the VAS anhedonia scale (difference between post-treatment and baseline mean: JNJ-67953964-1.27 cm; Placebo-1.01 cm) associated with an effect-size of 0.3, this was not statistically significant in this study, which was powered to detect relatively larger effect-sizes we anticipated for the primary imaging outcome measure. The relatively smaller effect-size seen with the VAS scale than the SHAPS and TEPS is surprising in light of the

history of VAS scales being relatively sensitive measures. However, the effect-size seen with the VAS anhedonia is consistent with the relatively smaller effect-sizes seen with the SHAPS and TEPS than the neuroimaging measures and further supports the hypothesis discussed above that the neuroimaging measures are likely to be associated with larger effects possibly because they are closer to the direct biological effects of the drug than the clinical measures.

Factors Related to Why We Did Not Find a Significant Effect for the Planned Analysis for the PRT Data:

There are a number of factors related to why we did not find a significant effect for the planned Treatment Arm x Block x Time interaction effect but did find a significant Treatment Arm x Time effect. In retrospect, our planned PRT analysis was based on a hypothesis of 3-way interaction involving Treatment Arm (KOR, Placebo), Block (block 1, block 2), and Time (pre-treatment, post-treatment) which we now believe was a suboptimal approach to analyzing our data. There are two reasons for this assessment. First, subject burden time limitations prevented us from implementing the 3-block version of the PRT, which has been used by over 50 groups worldwide in over 40 publications, and has been reliably found to induce systematic increases in response bias over the three blocks among healthy controls (typically, manifested as a main effect of Block for response bias). Instead, we used a 2-block version. Unfortunately, analyses of independent samples performed after we had decided on the analysis strategy for this study show that response bias does not increase as

much across blocks in the 2-block version of the PRT as in the 3-block version, thereby decreasing the chances that we would find a significant effect on the planned Treatment Arm by Block by Time Interaction.[4] Second, prior studies using the PRT in MDD had found a main effect of Group (rather than a Group x Block interaction), due to overall (i.e., averaged across blocks) response bias in MDD patients relative to healthy controls[2] and differences in response bias have been reported to differentiate depressed patients and those with severe anhedonia from healthy controls.[5-8] In retrospect, this would have been a more appropriate choice of planned analysis than the 3-way interaction.

REFERENCES FOR SUPPLEMENT

1. Braun J, Held L, Ledergerber B; Swiss HIV Cohort Study. Accounting for baseline differences and measurement error in the analysis of change over time. Stat Med. 33(1), 2-16 (2014).
2. Harrison L, Dunn DT, Green H, Copas AJ. Modelling the association between patient characteristics and the change over time in a disease measure using observational cohort data Stat Med. 28(26), 3260-75 (2009).
3. European Medicines Agency. Guideline on Adjustment for Baseline Covariates in Clinical Trials. (2015).

4. Webb CA, Dillon DG, Pechtel P, Goer FK, Murray L, Huys QJ, Fava M, McGrath PJ, Weissman M, Parsey R, Kurian BT, Adams P, Weyandt S, Trombello JM, Grannemann B, Cooper CM, Deldin P, Tenke C, Trivedi M, Bruder G, Pizzagalli DA. Neural correlates of three promising endophenotypes of depression: Evidence from the EMBARC study. Neuropsychopharmacology. 41(2), 454-63 (2016)
5. Pizzagalli DA, Iosifescu D, Hallett LA, Ratner KG, Fava M. Reduced hedonic capacity in major depressive disorder: evidence from a probabilistic reward task. J Psychiatr Res. 43(1), 76-87 (2008).
6. Pizzagalli DA, Jahn AL, O'Shea JP. Toward an objective characterization of an anhedonic phenotype: a signal-detection approach. Biol Psychiatry. 57(4), 319-27 (2005).
7. Fletcher K, Parker G, Paterson A, Fava M, Iosifescu D, Pizzagalli DA. Anhedonia in melancholic and non-melancholic depressive disorders. J Affect Disord. 184, 81-8 (2015).
8. Carter RM, McInnes JJ, Huettel SA, Adcock RA. Activation in the VTA and Nucleus Accumbens increases in anticipation of both gains and losses. Frontiers in Behav Neurosci. 3,1-15 (2009).
9. Olejnik S and Algina J (2003). Generalized Eta and Omega Squared Statistics: Measures of effect size for some common research designs. Psychological Methods 8(4) 434-447.

10. Okada K. Negative estimate of variance-accounted-for effect size: How often it is obtained, and what happens if it is treated as zero Behavior Research Methods. June 2017, Volume 49, Issue 3, pp 979–987|

Supplementary Table 1
Baseline Characteristics of the Analysis Sub-Cohorts: All Subjects

Variable	ITT Population N=89	As Treated Population N=86	Per Protocol Population N=86	Completers Population N=68
Mean Age in Years (SD)	39.5 (13.2)	39.5 (13.0)	39.5 (13.0)	40.0 (13.6)
Gender - %Female	62.9	62.8	62.8	58.8
Race				
%Caucasian	67.8	67.9	67.9	67.2
%African American	20.7	21.4	21.4	22.4
%Asian	3.4	2.4	2.4	3.0
%American Indian/Alaskan Native	1.1	1.2	1.2	0.0
%More Than One Race	6.9	7.1	7.1	7.5
Ethnicity - %Hispanic Origin	11.6	12.0	12.0	11.9
Mean BMI (SD)	28.7 (6.2)	28.9 (6.2)	28.9 (6.2)	29.1 (6.0)
Mean Weight (lbs) (SD)	180.6 (41.9)	182.2 (41.5)	182.2 (41.5)	184.6 (40.4)
Mean Baseline fMRI Ventral Striatal Activation in MID Task in Anticipation of **Gain** Contrasted with No-incentive Trials (SD)**	0.63 (0.8) (N=88)	0.63 (0.9) (N=85)	0.64 (0.8) (N=85)	0.57 (0.8) (N=67)
Mean Maximum Baseline fMRI Ventral Striatal Activation in MID Task (SD) in Anticipation of **Gain** Contrasted with No-Incentive Trials (SD)	2.70 (1.2) (N=88)	2.71 (1.2) (N=85)	2.71 (1.2) (N=85)	2.66 (1.1) (N=67)
Mean Baseline fMRI Ventral Striatal Activation in MID Task in Anticipation of **Loss** Contrasted with No-incentive Trials (SD)	0.33 (0.7) (N=88)	0.34 (0.7) (N=85)	0.34 (0.7) (N=85)	0.30 (0.7) (N=67)
Mean Maximum Baseline fMRI Ventral Striatal Activation in MID Task (SD) in Anticipation of **Loss** Contrasted with No-incentive Trials (SD)	2.19 (1.0) (N=88)	2.19 (1.0) (N=85)	2.19 (1.0) (N=85)	2.16 (1.0) (N=67)
Mean Baseline PRT Change in Response Bias from Block 1 to Block 2 (SD)*	0.04 (0.2) (N=76)	0.04 (0.2) (N=74)	0.04 (0.2) (N=74)	0.04 (0.2) (N=55)
Mean Baseline SHAPS (SD)*	34.9 (7.4) (N=88)	34.8 (7.5)	34.8 (7.5)	34.5 (6.8)
Mean Baseline PRT Response Bias (averaged across blocks) (SD)	0.11 (0.03) (N=76)	0.11 (0.03) (N=74)	0.11(0.03) (N=74)	0.12 (0.1) (N=55)
Mean Baseline EEfRT (SD)	0.36 (0.2) (N=83)	0.37 (0.2) (N=81)	0.37 (0.2) (N=81)	0.37 (0.2) (N=63)
Mean Baseline TEPS Anticipatory Subscore (SD)	29.4 (5.7) (N=88)	29.4 (5.7)	29.4 (5.7)	29.6 (5.9)
Means Baseline TEPS Consummatory Subscore (SD)	26.2 (4.4) (N=88)	26.3 (4.5)	26.3 (4.5)	26.3 (4.6)
Mean Baseline VAS Anhedonia (SD)	3.26 (2.2) (N=88)	3.24 (2.2)	3.24 (2.2)	3.25 (2.1)
Mean Baseline Resting State EEG Delta Current Density in Rostral Anterior Cingulate (SD)	74.0 (78.6) (N=81)	74.8 (79.4) (N=79)	74.8 (79.4) (N=79)	77.7 (84.6) (N=64)
Mean Baseline HAM-D (SD)	15.6 (5.6)	15.4 (5.6)	15.4 (5.6)	14.9 (5.3)
Mean Baseline HAM-A (SD)	15.5 (6.2)	15.5 (6.2)	15.5 (6.2)	14.9 (6.2)
Mean Baseline CGI-S (SD)	3.9 (0.5)	3.9 (0.6)	3.9 (0.6)	3.9 (0.5)
Mean Baseline CPFQ (SD)	26.3 (6.1)	26.2 (6.1)	26.2 (6.1)	25.4 (5.8)

Note: When N's are less than at top of column it reflects missing data for that variable

Supplementary Table 2: Baseline Characteristics of the JNJ-67953964 and Placebo Groups in Analysis Sub-Cohorts

Variable	ITT Population		As Treated Population		Per Protocol Population		Completers Population	
	JNJ N=45	Placebo N=44	JNJ N=43	Placebo N=43	JNJ N=43	Placebo N=43	JNJ N=34	Placebo N=34
Mean Age in Years (SD)	40.7 (13.3)	38.2 (13.0)	41.3 (13.0)	37.8 (12.9)	41.3 (13.0)	37.8 (12.9)	40.3 (13.8)	39.8 (13.5)
Gender - %Female	64.4	61.4	65.1	60.5	65.1	60.5	61.8	55.9
Race								
%Caucasian	70.5	65.1	71.4	64.3	71.4	64.3	70.6	63.6
%African American	22.7	18.6	23.8	19.0	23.8	19.0	23.5	21.2
%Asian	2.3	4.7	0.0	4.8	0.0	4.8	0.0	6.1
%American Indian/Alaskan Native	0.0	2.3	0.0	2.4	0.0	2.4	0.0	0.0
%More Than One Race	4.5	9.3	4.8	9.5	4.8	9.5	5.9	9.1
Ethnicity - %Hispanic Origin	11.6	11.6	12.2	11.9	12.2	11.9	11.8	12.1
Mean BMI (SD)	29.4 (6.4)	28.0 (5.9)	29.7 (6.4)	28.0 (6.0)	29.7 (6.4)	28.0 (6.0)	29.9 (6.1)	28.4 (5.9)
Mean Weight (lbs) (SD)	180.9 (43.7)	180.3 (40.6)	184.2 (42.3)	180.2 (41.1)	184.2 (42.3)	180.2 (41.1)	184.5 (39.8)	184.7 (41.5)
Mean Baseline fMRI Ventral Striatal Activation in MID Task in Anticipation of **Gain** Contrasted with No-incentive Trials (SD)**	0.63 (0.9) (N=44)	0.64 (0.8) (N=44)	0.63 (0.9) (N=42)	0.64 (0.8) (N=43)	0.63 (0.9) (N=42)	0.64 (0.8) (N=43)	0.58 (0.9) (N=33)	0.57 (0.7) (N=34)
Mean Maximum Baseline fMRI Ventral Striatal Activation in MID Task (SD) in Anticipation of **Gain** Contrasted with No-Incentive Trials (SD)	2.66 (1.2) (N=44)	2.73 (1.2) (N=44)	2.71 (1.2) (N=42)	2.71 (1.2) (N=43)	2.71 (1.2) (N=42)	2.71 (1.2) (N=43)	2.69 (1.2) (N=33)	2.63 (1.1) (N=34)
Mean Baseline fMRI Ventral Striatal Activation in MID Task in Anticipation of **Loss** Contrasted with No-incentive Trials (SD)	0.29 (0.8) (N=44)	0.36 (0.7) (N=44)	0.32 (0.8) (N=42)	0.36 (0.7) (N=43)	0.32 (0.8) (N=42)	0.36 (0.7) (N=43)	0.30 (0.8) (N=33)	0.30 (0.6) (N=34)
Mean Maximum Baseline fMRI Ventral Striatal Activation in MID Task (SD) in Anticipation of **Loss** Contrasted with No-incentive Trials (SD)	2.15 (1.2) (N=44)	2.23 (0.9) (N=44)	2.16 (1.1) (N=42)	2.21 (0.9) (N=43)	2.16 (1.1) (N=42)	2.21 (0.9) (N=43)	2.20 (1.2) (N=33)	2.13 (0.8) (N=34)
Mean Baseline PRT Change in Response Bias from Block 1 to Block 2 (SD)*	0.02 (0.2) (N=35)	0.05 (0.2) (N=41)	0.02 (0.2) (N=33)	0.05 (0.2) (N=41)	0.02 (0.2) (N=33)	0.05 (0.2) (N=41)	0.05 (0.2) (N=24)	0.03 (0.2) (N=31)
Mean Baseline SHAPS (SD)*	36.4 (8.5) (N=44)	33.4 (5.9) (N=44)	36.4 (8.6)	33.3 (5.9)	36.4 (8.6)	33.3 (5.9)	35.4 (8.1)	33.6 (5.2)
Mean Baseline PRT Response Bias (averaged across blocks) (SD)	0.11 (0.03) (N=35)	0.11 (0.03) (N=41)	0.11 (0.03) (N=33)	0.11 (0.03) (N=41)	0.11 (0.03) (N=33)	0.11 (0.03) (N=41)	0.11 (0.1) (N=24)	0.11 (0.1) (N=31)
Mean Baseline EEfRT (SD)	0.35 (0.2) (N=42)	0.38 (0.2) (N=41)	0.35 (0.2) (N=41)	0.38 (0.2) (N=40)	0.35 (0.2) (N=41)	0.38 (0.2) (N=40)	0.36 (0.2) (N=32)	0.38 (0.2) (N=31)
Mean Baseline TEPS Anticipatory Subscore (SD)	29.3 (5.7) (N=44)	29.5 (5.6) (N=44)	29.4 (5.7)	29.4 (5.7)	29.4 (5.7)	29.4 (5.7)	29.5 (5.9)	29.6 (5.8)
Means Baseline TEPS Consummatory Subscore (SD)	26.3 (4.4)	26.1 (4.4) (N=44)	26.4 (4.5)	26.1 (4.5)	26.4 (4.5)	26.1 (4.5)	26.4 (4.6)	26.1 (4.6)

	(N=44)							
Mean Baseline VAS Anhedonia (SD)	2.93 (2.1) (N=44)	3.59 (2.2) (N=44)	2.86 (2.1)	3.63 (2.2)	2.86 (2.1)	3.63 (2.2)	3.00 (2.2)	3.50 (2.0)
Mean Baseline Resting State EEG Delta Current Density in Rostral Anterior Cingulate (SD)	73.0 (97.1) (N=43)	75.2 (51.6) (N=38)	73.8 (98.0) (N=42)	76.0 (52.0) (N=37)	73.8 (98.0) (N=42)	76.0 (52.0) (N=37)	75.0 (104.7) (N=34)	80.7 (55.3) (N=30)
Mean Baseline HAM-D (SD)	16.3 (5.2)	14.8 (5.9)	16.0 (5.2)	14.8 (6.0)	16.0 (5.2)	14.8 (6.0)	14.7 (4.7)	15.0 (6.0)
Mean Baseline HAM-A (SD)	16.0 (5.8)	15.1 (6.6)	15.8 (5.7)	15.1 (6.7)	15.8 (5.7)	15.1 (6.7)	14.2 (5.2)	15.6 (7.2)
Mean Baseline CGI-S (SD)	3.9 (0.6)	4.0 (0.5)	3.9 (0.6)	4.0 (0.5)	3.9 (0.6)	4.0 (0.5)	3.8 (0.5)	4.0 (0.5)
Mean Baseline CPFQ (SD)	27.2 (6.4) (N=44)	25.4 (5.7) (N=44)	27.1 (6.4)	25.4 (5.8)	27.1 (6.4)	25.4 (5.8)	25.9 (5.9)	24.9 (5.7)

*JNJ = JNJ-67953964; Note: When N's are less than at top of column it reflects missing data for that variable

Supplementary Table 3. Site Effects on Outcomes Variables (ITT Population)

Variable	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site Effect*	SitexTime Effect*	SiteXArmXTime Effect*
Mean fMRI Ventral Striatal Activation in MID Task in Anticipation of Gain Contrasted with No-incentive Trials	JNJ N=1 0.99 (N/A) PCBO N=5 0.13 (0.72)	JNJ N=11 0.99 (0.73) PCBO N=9 0.78 (0.81)	JNJ N=2 0.24 (1.0) PCBO N=5 0.11 (0.96)	JNJ N=12 0.77 (0.87) PCBO N=9 0.00 (0.81)	JNJ N=9 0.59 (0.90) PCBO N=9 0.35 (0.90)	JNJ N=9 0.60 (0.81) PCBO N=7 0.41 (0.82)	p=0.19 η^2=0.02 ω^2=0.00	p=0.47 η^2=0.01 ω^2=0.0	p=0.45 η^2=0.01 ω^2=-0.01
Mean SHAPS	JNJ N=1 30.1 (N/A) PCBO N=5 30.8 (4.7)	JNJ N=11 28.3 (5.0) PCBO N=9 33.6 (4.8)	JNJ N=2 35.9 (4.9) PCBO N=5 35.8 (4.7)	JNJ N=12 29.4 (4.5) PCBO N=9 30.8 (4.8)	JNJ N=9 31.1 (4.8) PCBO N=9 33.8 (5.1)	JNJ N=9 34.9 (5.4) PCBO N=7 32.1 (4.8)	p=0.026 η^2=0.04 ω^2=0.03	p=0.47 η^2=0.01 ω^2=0.00	P=0.016 η^2=0.04 ω^2=-0.02
Mean PRT Change in Response Bias from Block 1 to Block 2	JNJ N=1 0.40 (N/A) PCBO N=6 0.08 (0.16)	JNJ N=9 -0.10 (0.16) PCBO N=7 0.00 (0.17)	JNJ N=1 0.01 (N/A) PCBO N=5 0.11 (0.16)	JNJ N=10 0.06 (0.18) PCBO N=9 0.07 (0.17)	JNJ N=9 -0.02 (0.16) PCBO N=8 0.13 (0.17)	JNJ N=5 -0.03 (0.16) PCBO N=6 0.22 (0.17)	p=0.35 η^2=0.02 ω^2=-0.01	p=0.055 η^2=0.08 ω^2=-0.03	p=0.11 η^2=0.12 ω^2=0.06
Maximum fMRI Ventral Striatal Activation in MID Task in Anticipation of **Gain** Contrasted with No-incentive Trials	JNJ N=1 4.5 (N/A) PCBO N=5 1.8 (0.89)	JNJ N=11 3.0 (0.93) PCBO N=9 2.3 (0.90)	JNJ N=2 2.6 (1.2) PCBO N=5 1.4 (1.1)	JNJ N=12 3.0 (1.1) PCBO N=9 2.2 (0.90)	JNJ N=9 2.7 (1.1) PCBO N=9 3.1 (1.1)	JNJ N=9 2.6 (0.99) PCBO N=7 2.6 (1.0)	p=0.48 η^2=0.02 ω^2=0.00	p=0.27 η^2=0.01 ω^2=0.00	p=0.26 η^2=0.03 ω^2=0.01
Mean fMRI Ventral Striatal Activation in MID Task in Anticipation of **Loss** Contrasted with No-incentive Trials	JNJ N=1 1.9 (N/A) PCBO N=5 0.09 (0.40)	JNJ N=11 0.94 (0.66) PCBO N=9 0.36 (0.60)	JNJ N=2 -0.2 (0.85) PCBO N=5 0.52 (0.80)	JNJ N=12 0.68 (0.69) PCBO N=9 -0.49 (0.69)	JNJ N=9 0.54 (0.75) PCBO N=9 0.24 (0.78)	JNJ N=9 0.64 (0.69) PCBO N=7 0.01 (0.71)	p=0.056 η^2=0.02 ω^2=0.00	p=0.21 η^2=0.01 ω^2=0.00	p=0.11 η^2=0.03 ω^2=0.01
Maximum fMRI Ventral Striatal Activation in MID Task in Anticipation of **Loss** Contrasted with No-incentive Trials	JNJ N=1 4.4 (N/A) PCBO N=5 2.0 (0.60)	JNJ N=11 3.1 (0.93) PCBO N=9 2.3 (0.99)	JNJ N=2 1.5 (1.2) PCBO N=5 2.2 (1.2)	JNJ N=12 2.7 (1.1) PCBO N=9 2.1 (0.99)	JNJ N=9 2.5 (1.1) PCBO N=9 2.8 (1.1)	JNJ N=9 2.5 (0.99) PCBO N=7 1.8 (1.0)	p=0.25 η^2=0.01 ω^2=0.00	p=0.34 η^2=0.01 ω^2=0.00	p=0.23 η^2=0.02 ω^2=0.00
Mean Baseline PRT Response Bias (averaged across blocks)	JNJ N=1 0.17 (N/A) PCBO N=6 0.08 (0.14)	JNJ N=9 0.17 (0.13) PCBO N=7 0.09 (0.12)	JNJ N=1 0.18 (N/A) PCBO N=5 0.25 (0.13)	JNJ N=10 0.19 (0.15) PCBO N=9 0.03 (0.14)	JNJ N=9 0.18 (0.13) PCBO N=8 0.01 (0.15)	JNJ N=5 0.05 (0.14) PCBO N=6 0.00 (0.14)	p=0.02 η^2=0.09 ω^2=0.05	p=0.15 η^2=0.08 ω^2=0.04	p=0.41 η^2=0.08 ω^2=-0.01

TEPS Anticipatory Subscale	JNJ N=1 41.2 (N/A)	JNJ N=11 37.4 (6.0)	JNJ N=2 18.0 (7.5)	JNJ N=12 31.5 (5.5)	JNJ N=9 32.1 (6.3)	JNJ N=9 28.4 (6.3)	p=0.008 η^2=0.04 ω^2=0.03	p=0.035 η^2=0.01 ω^2=0.00	P=0.001 η^2=0.04 ω^2=0.03
	PCBO N=5 31.3 (5.6)	PCBO N=9 35.7 (6.3)	PCBO N=5 30.3 (5.6)	PCBO N=9 32.9 (5.7)	PCBO N=9 29.2 (6.0)	PCBO N=7 32.7 (6.3)			
TEPS Consummatory Subscale	JNJ N=1 27.1 (N/A)	JNJ N=11 31.5 (4.6)	JNJ N=2 34.6 (5.9)	JNJ N=12 28.9 (4.5)	JNJ N=9 31.2 (5.1)	JNJ N=9 24.5 (5.1)	p=0.09 η^2=0.02 ω^2=0.01	p=0.075 η^2=0.01 ω^2=0.00	P=0.35 η^2=0.00 ω^2=0.00
	PCBO N=5 30.2 (4.5)	PCBO N=9 28.1 (5.1)	PCBO N=5 26.2 (4.4)	PCBO N=9 28.8 (4.5)	PCBO N=9 24.0 (4.8)	PCBO N=7 25.5 (5.0)			
EEfRT	JNJ N=1 0.22 (N/A)	JNJ N=10 0.32 (0.13)	JNJ N=2 .45 (0.17)	JNJ N=11 0.49 (0.14)	JNJ N=9 0.34 (0.15)	JNJ N=9 0.52 (0.15)	p=0.055 η^2=0.01 ω^2=0.00	p=0.022 η^2=0.03 ω^2=0.02	P=0.42 η^2=0.01 ω^2=0.00
	PCBO N=4 0.36 (0.16)	PCBO N=8 0.41 (0.12)	PCBO N=5 0.45 (0.18)	PCBO N=8 0.47 (0.15)	PCBO N=9 0.34 (0.15)	PCBO N=7 0.42 (0.13)			
VAS Anhedonia	JNJ N=1 5.4 (N/A)	JNJ N=11 5.2 (1.7)	JNJ N=2 1.4 (2.3)	JNJ N=12 3.6 (1.7)	JNJ N=9 4.7 (1.8)	JNJ N=9 3.3 (1.8)	p=0.055 η^2=0.02 ω^2=0.01	p=0.022 η^2=0.02 ω^2=0.00	P=0.42 η^2=0.02 ω^2=0.00
	PCBO N=5 4.8 (1.6)	PCBO N=9 4.8 (1.8)	PCBO N=5 4.2 (1.8)	PCBO N=9 4.7 (1.8)	PCBO N=9 4.0 (1.8)	PCBO N=7 4.5 (1.9)			
Resting State EEG Delta Current Density in Rostral Anterior Cingulate	JNJ N=1 15.2 (N/A)	JNJ N=10 72.5 (84.7)	JNJ N=4 27.6 (160.1)	JNJ N=11 50.3 (88.9)	JNJ N=9 51.1 (115.5)	JNJ N=8 79.7 (107.2)	p=0.35 η^2=0.04 ω^2=0.02	p=0.48 η^2=0.03 ω^2=0.01	P=0.26 η^2=0.02 ω^2=0.00
	PCBO N=5 220.2 (75.8)	PCBO N=8 55.0 (99.8)	PCBO N=4 76.5 (86.8)	PCBO N=9 44.6 (100.8)	PCBO N=7 25.5 (78.0)	PCBO N=5 76.9 (84.5)			
HAM-D	JNJ N=2 4.7 (5.1)	JNJ N=11 8.7 (4.3)	JNJ N=2 12.9 (5.5)	JNJ N=12 10.7 (4.5)	JNJ N=9 10.3 (4.8)	JNJ N=9 14.5 (4.5)	p=0.09 η^2=0.01 ω^2=0.00	p=0.06 η^2=0.02 ω^2=0.01	P=0.31 η^2=0.01 ω^2=0.00
	PCBO N=5 10.8 (4.0)	PCBO N=9 9.2 (4.5)	PCBO N=5 12.9 (4.4)	PCBO N=9 8.2 (4.5)	PCBO N=9 14.2 (4.8)	PCBO N=7 12.2 (4.8)			
HAM-A	JNJ N=2 2.0 (5.4)	JNJ N=11 11.3 (4.6)	JNJ N=2 22.3 (5.9)	JNJ N=12 10.5 (4.8)	JNJ N=9 8.6 (5.1)	JNJ N=9 13.1 (4.8)	p=0.03 η^2=0.01 ω^2=0.00	p=0.02 η^2=0.02 ω^2=0.01	P=0.17 η^2=0.01 ω^2=0.00
	PCBO N=5 8.8 (4.2)	PCBO N=9 10.2 (4.8)	PCBO N=5 14.0 (4.7)	PCBO N=9 7.4 (4.8)	PCBO N=9 12.7 (4.8)	PCBO N=7 11.7 (5.0)			
CGI-I	JNJ N=2 3.8 (1.2)	JNJ N=11 3.1 (1.0)	JNJ N=2 3.0 (1.1)	JNJ N=12 3.5 (1.0)	JNJ N=9 2.8 (0.9)	JNJ N=9 3.6 (0.9)	p=0.45 η^2=0.02 ω^2=0.01	p=0.06 η^2=0.03 ω^2=0.00	P=0.012 η^2=0.06 ω^2=0.02

	PCBO N=5 3.1 (0.9)	PCBO N=9 3.2 (0.9)	PCBO N=5 2.9 (0.9)	PCBO N=9 2.9 (0.9)	PCBO N=9 3.7 (0.9)	PCBO N=7 3.1 (1.1)			
CGI-S	JNJ N=2 2.8 (1.0)	JNJ N=11 3.1 (0.7)	JNJ N=2 3.5 (0.9)	JNJ N=12 3.3 (0.7)	JNJ N=9 2.8 (0.9)	JNJ N=9 3.4 (0.9)	p=0.35 η^2=0.01 ω^2=0.00	p=0.32 η^2=0.02 ω^2=0.00	P=0.43 η^2=0.02 ω^2=0.00
	PCBO N=5 3.3 (0.7)	PCBO N=9 3.1 (0.6)	PCBO N=5 3.4 (0.7)	PCBO N=9 3.0 (0.6)	PCBO N=9 3.3 (0.9)	PCBO N=7 3.4 (0.8)			
CPFQ	JNJ N=1 17.6 (N/A)	JNJ N=11 19.7 (4.6)	JNJ N=2 24.5 (5.8)	JNJ N=12 22.5 (4.8)	JNJ N=9 17.8 (5.1)	JNJ N=9 24.0 (4.5)	p=0.20 η^2=0.01 ω^2=0.00	p=0.27 η^2=0.01 ω^2=0.00	P=0.075 η^2=0.02 ω^2=0.01
	PCBO N=5 18.7 (4.0)	PCBO N=9 23.0 (4.5)	PCBO N=5 23.8 (4.5)	PCBO N=9 17.4 (4.5)	PCBO N=9 23.0 (4.8)	PCBO N=7 21.5 (4.8)			

Site columns contain baseline corrected least squared means (SD) at end of double-blind treatment from mixed effects models;

* η^2 and ω^2 are measures of effect-size for ANOVA effects. ω^2 is a relatively unbiased .estimate for effect-size for ANOVA compared with η^2 and can be negative with a possible range from -1 to 1. Negative values occur when F is less than 1.[9,10]