

ORIGINAL ARTICLE

Central monitoring of depression and anxiety symptoms reduces placebo responses in depression clinical trials: A post hoc exploratory analysis of data from the phase III CCT-004 trial of vortioxetine

Yoshinori Watanabe¹ | Akira Nishimura²  | Toshiaki Kikuchi³ | Norifusa Sawada⁴ | Manami Imazaki⁵ | Isao Inada⁵ | Koichiro Watanabe⁶

¹Himorogi Psychiatric Institute, Tokyo, Japan

²Pharmaceutical Development Division, Towa Pharmaceutical Co., Ltd., Osaka, Japan

³Department of Neuropsychiatry, Keio University School of Medicine, Tokyo, Japan

⁴Mental and Sleep Disorder Clinic of Narimasu, Tokyo, Japan

⁵Neuroscience Therapeutic Area Strategy Unit, Takeda Development Center Japan, Takeda Pharmaceutical Company Limited, Osaka, Japan

⁶Department of Neuropsychiatry, Kyorin University School of Medicine, Tokyo, Japan

Correspondence

Akira Nishimura, MBA, Pharmaceutical Development Division, Towa Pharmaceutical Co., Ltd. 2-11 Shinbashi-cho, Kadoma, Osaka 571-8580, Japan.
Email: a-nishimura@towayakuhin.co.jp

Funding information

Takeda Pharmaceutical Co., Ltd.

Abstract

Aim: Clinical trials of antidepressants often fail to demonstrate their efficacy versus placebo, suggesting that patient selection based on physician ratings of depression may contribute to a high placebo response.

Methods: In the CCT-004 trial of vortioxetine, central monitoring was employed to compare physician and patient ratings of depression and anxiety at baseline and over time to identify factors contributing to a large placebo response, as well as to explore the potential of a unique patient-rated clinical measure combining QIDS-J and Himorogi Self-rating Anxiety Scale (HSAS), to contribute to optimal patient selection at baseline and patient monitoring over time.

Results: The CCT-004 trial showed similar trends between the QIDS-J and MADRS (Montgomery-Åsberg Depression Rating Scale) ratings. It was suggested that central monitoring of the QIDS-J and MADRS ratings of depression and anxiety symptoms helped reduce the baseline score inflation by calling the study sites' attention to discrepancies between these ratings at baseline; it also allowed these ratings to be assessed for their concordance over time. Of note, MDD patients with baseline QIDS-J scores ≥ 11 /HSAS ≤ 19 were associated with the smallest placebo response, with the effect size being larger than that for those with QIDS-J scores ≤ 10 /HSAS ≥ 20 .

Conclusion: The use of both physician and patient ratings of depression and anxiety symptoms at baseline and over time, as well as their central monitoring, helped minimize the baseline score inflation and optimize patient monitoring over time, and allowed the antidepressant to be evaluated for its full therapeutic potential.

KEYWORDS

anxiety, drug monitoring, major depressive disorder, placebo effect, randomized controlled trials

(Trial registration: [ClinicalTrials.gov](https://clinicaltrials.gov) identifier: NCT02389816)

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2022 The Authors. *Neuropsychopharmacology Reports* published by John Wiley & Sons Australia, Ltd on behalf of The Japanese Society of Neuropsychopharmacology.



1 | INTRODUCTION

Patients with major depressive disorder (MDD) are often associated with a large placebo response, which translates into a small effect size for antidepressants, thus leading to the failure of clinical trials to confirm their efficacy over placebo,¹ while the placebo response is shown to increase over time in placebo-controlled trials.²

Indeed, despite their regulatory approval overseas, some antidepressants, such as fluoxetine, remain unapproved in Japan, primarily due to the failure of domestic clinical trials to demonstrate their efficacy versus placebo.

In this connection, research suggests that patients with severe MDD respond well to antidepressants and poorly to placebo; however, patients with mild MDD respond equally well to antidepressants and placebo,³ thus making it difficult to evaluate antidepressants for their full therapeutic potential or their effect size, while at the same time highlighting the therapeutic role of placebo in the treatment of those with mild-to-moderate MDD.⁴

Attention has recently been focused not only on reducing placebo responses through modification of study designs, which, however, appears to be unlikely to increase the treatment effect,⁵ but on reducing baseline rater bias or score inflation, which is thought likely to contribute to a large placebo response through inclusion of those with mild to moderate depression who are shown to be associated with a larger placebo than antidepressant response.

Of note, a potential role has been suggested for centralized rating, as opposed to site-based rating, in facilitating selection of candidates for trials of MDD, as well as in reducing placebo responses.⁶

Despite a paucity of evidence that patient rating improves treatment outcomes in those with MDD,⁷ recent research suggests a potential role for patient ratings of depression severity in clinical trials.⁸⁻¹⁰

Against this background, an exploratory analysis has been performed on the data available mainly from the phase III CCT-004 trial of the antidepressant vortioxetine, which employed central monitoring to compare physician and patient ratings of severity of both depression and anxiety symptoms in an attempt to provide insight into patient selection at baseline and patient monitoring over time using relevant clinical measures, which may prove helpful in facilitating not only patient selection but also signal detection in clinical trials of MDD.

2 | METHODS

This was a post hoc analysis of a multicenter, randomized, double-blind, placebo-controlled, phase III (CCT-004) trial conducted in Japan from April 10, 2015, to March 16, 2018, to investigate the efficacy and safety of vortioxetine in Japanese patients with MDD.¹¹ The protocol of the study was approved by the Institutional Review

Board (IRB) of Takeda Pharmaceutical Co., Ltd. on September 1, 2014, and was conducted in accordance with the principles defined by the IRB, Good Clinical Practices guidelines and all other applicable regulatory requirements. All patients provided written informed consent prior to study enrollment. All investigations were performed in accordance with the Declaration of Helsinki and International Council on Harmonization tripartite guideline on the ethical principles of Good Clinical Practices.

The CCT-004 trial investigated the safety and efficacy of 8-week treatment with vortioxetine 10 mg ($n = 165$) and 20 mg ($n = 163$) in Japanese adults with recurrent MDD,¹¹ based on efficacy results from the preceding two short-term trials (CCT-002 and CCT-003) with the primary endpoint for the study being the change from baseline in Montgomery-Åsberg Depression Rating Scale (MADRS) total score at week 8 following initiation of the study treatment.

The inclusion and exclusion criteria in the CCT-004 trial, as well as its study design, have previously been reported in detail.¹¹ Briefly, the CCT-004 trial enrolled patients with a primary diagnosis of recurrent MDD defined according to the DSM-IV-TR criteria, whose current episode lasted 3–12 months and who required to have an MADRS total score of ≥ 26 , a HAM-17 total score of ≥ 18 , and a Clinical Global Impression of Severity (CGI-S) score of ≥ 4 throughout the screening and placebo run-in periods and at entry into the double-blind phase of the trial.

In addition to physician ratings, that is, MADRS, used for primary efficacy assessments in view of the new drug application for regulatory review and approval, the study also involved patient ratings of depression and anxiety using re-QIDS-J as an adjunct, which combined QIDS-J, a Japanese self-report version of the Inventory of Depression Symptomatology (IDS) developed to address the limitations of HAM-D17,¹² with the Himorogi Self-rating Anxiety Scale (HSAS),¹³ a Japanese self-rating anxiety scale developed to assess the severity of anxiety symptoms which were thought likely to affect the efficacy assessment of the study medication versus placebo.¹⁴

2.1 | Post hoc analysis

This post hoc analysis mainly included all patients randomized to placebo and vortioxetine in the CCT-004 which involved central monitoring, unlike the phase II CCT-002 and CCT-003 trials, but also those randomized to placebo and vortioxetine in the two-phase II trials to compare trends in placebo response over time between the CCT-002, CCT-003, and CCT-004 trials.

In this exploratory post hoc analysis, changes from baseline in MADRS total score at week 8 were examined among those treated with placebo and vortioxetine 10 and 20 mg in the CCT-002, CCT-003, and CCT-004 trials to compare responses to placebo versus the active treatment.

Also, all clinical rating scale scores were examined to identify factors likely contributing to placebo responses.



In addition, with a focus on the CCT-004 trial which involved central monitoring, central monitoring was examined for its role in mediating between physician and patient ratings of anxiety and depression at baseline and over time and thus in optimizing patient selection and monitoring. When any clear dissociation between these ratings was detected by the members of the Assessment Monitoring Committee during the trial, the site monitor (i.e., clinical research associate [CRA]) was called on to visit the site, sound the investigator's opinion about the dissociation by showing the QIDS-J and MADRS time series plots and provide feedback to the members of the Assessment Monitoring Committee. However, this did not involve changing the investigator's judgment, as this monitoring scheme was primarily intended to call attention to discrepancies between the patient and physician ratings, thereby promoting awareness that the investigators needed to exercise care in evaluating their subjects.

Finally, baseline patient characteristics were explored to identify patient profiles likely contributing to minimization of placebo responses over time in a clinical trial.

2.2 | Statistical analysis

The primary efficacy analysis¹¹ was performed to compare MADRS total scores at week 8 in the full analysis set (FAS) (i.e., all subjects randomized to receive ≥ 1 dose of the study medication in the double-blind period of the trial) using a mixed-effect model repeated measure (MMRM) approach with the change from baseline in the MADRS total score as a dependent variable, and visit, treatment group, visit-by-treatment group interaction, and baseline MADRS total score-by-visit interaction as fixed effects.

An analysis of covariance (ANCOVA) model was used to analyze the change from baseline in MADRS total score at week 8 (LOCF) with treatment as a fixed effect and baseline MADRS total score as a covariate. A logistic regression model was used to compare odds ratios for MADRS response/remission at week 8 (LOCF) as a dependent variable, and treatment group and baseline MADRS total score as independent variables.

In analyzing individual subscale scores, descriptive statistics and two-sided 95% confidence intervals of means were provided for the observed values and single-item changes from baseline at each post-dose visit by treatment group.

2.3 | Post hoc analysis

In this post hoc analysis, Spearman's correlation coefficients were calculated by using the number of patients randomized to placebo in the CCT-002, CCT-003, and CCT-004 trials to assess correlation between the baseline MADRS subscales and the change from baseline in MADRS total score at week 8; and the CORRELL Excel function was used to assess correlation between the MADRS and QIDS-J scores at baseline and at week 8 in the CCT-004 trial.

Additionally, Cohen's *d* was used to calculate the effect size for MADRS as stratified by baseline re-QIDS-J score in the CCT-004 trial.

3 | RESULTS

The results of the CCT-004 trial have previously been reported in detail.¹¹ Briefly, of the 662 subjects who provided informed consent, 530 entered the placebo run-in phase, and 493 were randomized. Of these, 453 completed the double-blind phase. The study population had a mean age of 40 years (males, 54.6%) and a mean baseline MADRS total score of 30.6. In the CCT-004 trial, those given vortioxetine 10 and 20 mg showed a significantly higher MADRS total score at week 8 (-15.03 and -15.45 , respectively) than those given placebo (-12.37) and significant improvements in the secondary depression-related endpoints of HAM-D17 score, CGI-I score, and SDS total score at week 8, with no significant difference was shown in DSST total score, an objective measure of cognitive function, despite a significant improvement in individual perception of cognitive function as assessed by subjective, patient reported PDQ-5 score, with a similar trend observed across the PDQ-5 subscales.

In this post hoc analysis, all the data available from the CCT-002, CCT-003, and CCT-004 trials were drawn on to evaluate relevant clinical measures for their association with a placebo response as well as to explore the role of central monitoring in optimizing patient selection and monitoring of antidepressant efficacy in clinical trials. A summary of findings from this post hoc analysis is given below.

3.1 | Contributors to a placebo response in the CCT-002, CCT-003, and CCT-004 trials

A review of the data from the CCT-002, CCT-003, and CCT-004 trials (Table S1) consistently revealed no increase in placebo response over time (Table S2a–c). Again, of these trials, the CCT-004 trial was shown to be associated with the smallest placebo response. Furthermore, an examination of relevant subscales through central monitoring (Table 1) demonstrated that reduced appetite in MADRS (MADRS05) was most associated with an increased placebo response over time (Figure S1), while increased appetite in QIDS-J (QIDS-J07) was associated with a decreased placebo response (Figure 1; see also Table S3).

3.2 | Central monitoring of MADRS and QIDS-J ratings

In this study, central monitoring led to the study sites being alerted in a timely fashion to those patients whose QIDS-J and MADRS ratings widely differed at baseline and over time and thus helped optimize patient selection and monitoring in the CCT-004 trial. Thus, QIDS-J and MADRS ratings showed similar trends across all their corresponding subscales/domains (Figure 1).

3.3 | MADRS total scores as stratified by baseline QIDS-J/HSAS score

An analysis of MADRS total scores by baseline QIDS-J and HSAS scores showed that those with QIDS-J scores ≥ 11 /HSAS ≤ 19 were associated with a smaller placebo response than those with QIDS-J scores ≤ 10 and HSAS ≥ 20 (Table 2).

TABLE 1 Correlation between the baseline MADRS subscales and the change from baseline in MADRS total score

MADRS individual subscale (baseline)	Change from baseline in the MADRS total score at week 8 (LOCF)		
	n ^a	Correlation coefficient	P-value
Apparent sadness	434	-0.05979	0.2138
Reported sadness	434	-0.11769	0.0142
Inner tension	434	-0.05893	0.2205
Reduced sleep	434	0.00845	0.8607
Reduced appetite	434	-0.13259	0.0057
Concentration difficulties	434	-0.02752	0.5675
Lassitude	434	-0.06146	0.2013
Inability to feel	434	0.01873	0.6972
Pessimistic thoughts	434	-0.07005	0.1452
Suicidal thoughts	434	-0.06121	0.2031

Abbreviation: MADRS, Montgomery-Åsberg Depression Rating Scale.

^a Spearman's correlation coefficients were calculated by using the number of subjects randomized to placebo in the CCT-002 (n = 150), CCT-003 (n = 123) and CCT-004 (n = 161) to assess correlation between the baseline MADRS subscales and the change from baseline in MADRS total score.

Consistently, the effect size was shown to be larger in those with baseline QIDS-J scores ≥ 11 /HSAS ≤ 19 than in those with QIDS-J scores ≤ 10 and HSAS ≥ 20 (Table 3), with the effect size also shown to be enhanced over time for nearly all MADRS subscales (Table 4).

Furthermore, an analysis of the changes in MADRS total scores by baseline HSAS subscale revealed that those who had reported experiencing depersonalization "sometimes" and "frequently" at baseline were associated with a larger placebo response than those who had reported otherwise (Table S4).¹³

4 | DISCUSSION

This post hoc analysis showed that the CCT-004 trial was associated with the smallest placebo response of all three trials compared (CCT-002, CCT-003, and CCT-004). While this may have been influenced by the use of central monitoring, inclusion of a 1-week run-in period, and the focus on patients with recurrent MDD alone in the CCT-004 trial, it was suggested that the use of self-administered and objective measures of depression, as well as measures of anxiety that affects the assessment of MDD, may improve outcomes in clinical trials of antidepressants.

Indeed, the CCT-004 trial demonstrated that the effect size with vortioxetine was larger for those with baseline QIDS-J scores ≥ 11 /HSAS ≤ 19 than for those with baseline QIDS-J scores ≥ 10 /HSAS ≤ 20 , suggesting that those with baseline QIDS-J scores ≥ 11 /HSAS ≤ 19 may be targeted for inclusion in clinical trials of MDD to increase the probability of success in demonstrating the efficacy of antidepressants.

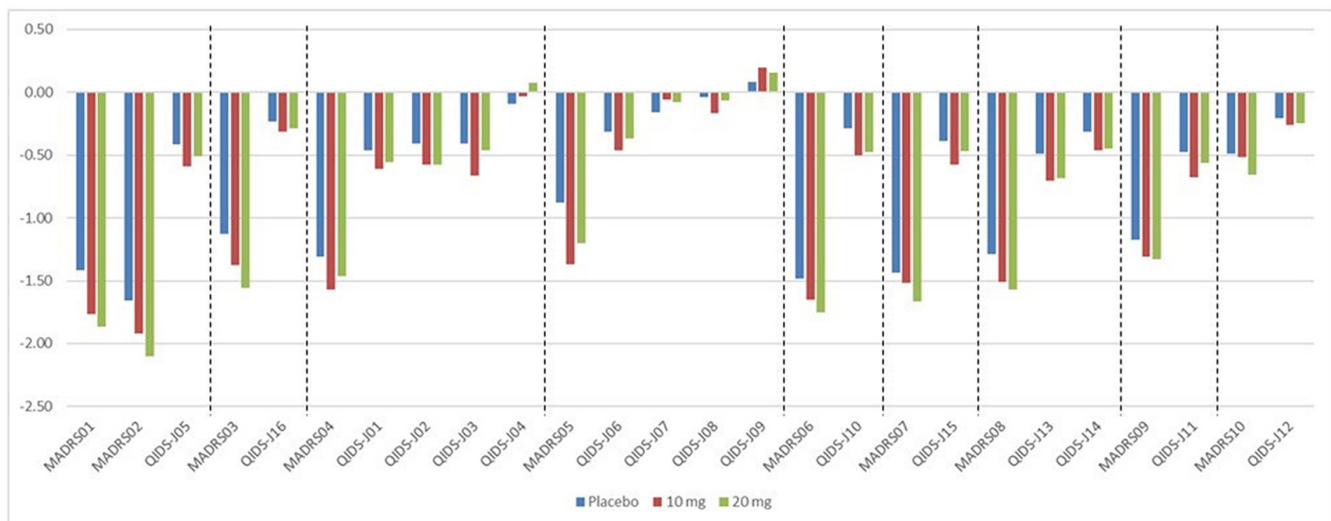


FIGURE 1 Change from baseline in the rQIDS-J/MADRS subscales, MADRS01, reported sadness; MADRS02, apparent sadness; MADRS03, inner tension; MADRS04, reduced sleep; MADRS05, reduced appetite; MADRS06, concentration difficulties; MADRS07, lassitude; MADRS08, inability to feel; MADRS09, pessimistic thoughts; MADRS10, suicidal thoughts; QIDS-J01, sleep-onset insomnia; QIDS-J02, mid-nocturnal insomnia; QIDS-J03, morning insomnia; QIDS-J04, hypersomnia; QIDS-J05, mood (sad); QIDS-J06, appetite (decrease); rQIDS-J07, appetite (increase); QIDS-J-8, weight (decrease); QIDS-J09, weight (increase); QIDS-J10, concentration/decision making; QIDS-J11, outlook (self); QIDS-J12, suicidal ideation; QIDS-J13, involvement; QIDS-J14, energy/fatigability; QIDS-J15, psychomotor slowing; QIDS-J16, psychomotor agitation; VOR, vortioxetine (See also Table S3)



TABLE 2 Changes from baseline in MADRS total scores at week 8 as stratified by baseline rQIDS-J/HSAS score

		Placebo	Vortioxetine 10 mg	Vortioxetine 20 mg		
All patients stratified by baseline rQIDS-J/HSAS score (A + B)	N	159	164	160		
	Baseline MADRS total score	30.57	30.79	30.62		
	Change in MADRS total score	-12.26	-14.52	-2.26	-15.15	-2.89
	Change in HSAS score	-2.72	-5.46	-2.74	-4.52	-1.80
	Change in rQIDS-J score	-3.16	-4.75	-4.17		
A Patients with baseline rQIDS-J \geq 11/ HSAS \leq 19	N	61	66	73		
	Baseline MADRS total score	30.74	30.37	30.74		
	Change in MADRS total score	-9.85	-14.24	-4.39	-14.77	-4.92
A Patients with baseline rQIDS \leq 10/ HSAS \geq 20	N	98	98	87		
	Baseline MADRS total score	30.46	31.03	30.51		
	Change in MADRS total score	-13.77	-14.71	-0.94	-15.48	-1.71

Abbreviations: MADRS, Montgomery-Åsberg Depression Rating Scale; rQIDS-J, revised Japanese version of the Quick Inventory of Depressive Symptomatology incorporating the Himorogi Self-rating Anxiety Scale (HSAS).

TABLE 3 Cohen's d (effect size) for MADRS total score as stratified by baseline re-QIDS-J score in the CCT-004 trial

		Placebo	Vortioxetine 10 mg	Vortioxetine 20 mg
Patients with baseline QIDS-J \geq 11/ HSAS \leq 19	N	61	66	73
	Baseline MADRS total score	30.74	30.37	30.74
	Change in MADRS total score	-9.85	-14.24	-14.77
	MADRS SE	9.38	8.89	8.82
	MADRS Cohen's d (effect size)	-	0.48	0.54
Patients with baseline QIDS-J \leq 10/ HSAS \geq 20	N	98	98	87
	Baseline MADRS total score	30.46	31.03	30.51
	Change in MADRS total score	-13.77	-14.71	-15.48
	MADRS Cohen's d (effect size)	-	0.11	0.19

Abbreviations: MADRS, Montgomery-Åsberg Depression Rating Scale; rQIDS-J, revised Japanese version of the Quick Inventory of Depressive Symptomatology incorporating the Himorogi Self-rating Anxiety Scale (HSAS).

This analysis also showed that the effect size increased over time for most MADRS subscales. Of all MADRS subscales, decreased appetite was most associated with the placebo response in the CCT-004 trial.

It was also shown that those who reported often experiencing depersonalization at baseline had a greater placebo response than those who reported otherwise. Although the effect of depersonalization symptoms on the placebo response requires further examination, it is suggested that depersonalization may be associated with the activation of the opioid system possibly involved in the placebo effect.¹⁵

In addition, there was no consistent increase in placebo response in the trials evaluated, contrary to earlier reports suggesting a tendency toward an increased placebo response over time in clinical trials.² This may have reflected the change in study design in the CCT-004 trial described above.

These findings have some important implications for patient selection and monitoring in clinical trials of MDD. First, while physician ratings, such as MADRS and HAM-D17, remain the gold standard for assessment of severity of MDD particularly due to the paucity of

evidence supporting the use of patient ratings, such as QIDS, in improving treatment outcomes in patients with MDD,⁷ recent research suggests high levels of agreement between patient and clinician ratings of baseline severity in treatment-resistant depression (TRD),⁸ the diagnostic validity of patient-rated severity of depression in the elderly,⁹ and the accuracy of both clinician and patient ratings of severity in children and adolescents,¹⁰ thus suggesting a potential role for patient rating of severity in clinical trials of MDD. Again, of note, the patient version of QIDS is shown to have the potential to prove as useful as the physician version of the QIDS or HAM-D17.¹⁶

Indeed, while the CCT-004 trial showed similar trends between the QIDS-J and MADRS ratings despite lack of correspondence between the MADRS and QIDS-J subscales in some domains (i.e., hypersomnia, appetite increase, weight decrease, and weight increase) (Table S2), study findings also suggested that the use of QIDS-J or QIDS in general might be well worth considering in clinical trials, given that its use helped clarify that "increased appetite" in QIDS-J was associated with an increased placebo response over time in the CCT-004 trial. In this regard, "reduced appetite" was shown to tend to improve, and "increased appetite" to remain almost unchanged,



TABLE 4 Cohen's d (effect size) for the MADRS subscales at week 8 as stratified by baseline re-QIDS-J score in the CCT-004 trial

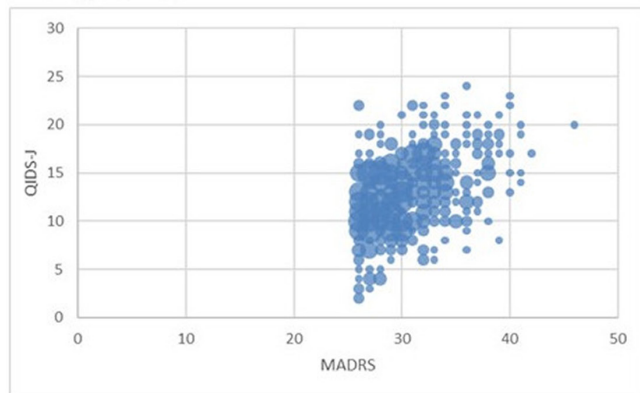
A	MADRS01	MADRS02	MADRS03	MADRS04	MADRS05	MADRS06	MADRS07	MADRS08	MADRS09	MADRS10	
Patients with baseline QIDS-J ≥ 11 /HSAS ≤ 19	Placebo (n = 61)	-1.16	-1.39	-0.95	-1.30	-0.62	-1.08	-1.11	-0.90	-0.41	
	Change										
	SE	1.37	1.55	1.16	1.50	1.49	1.36	1.28	1.26	1.22	
	Vortioxetine	-1.76	-1.95	-1.39	-1.59	-1.17	-1.70	-1.55	-1.39	-1.24	-0.50
	10 mg (n = 66)	1.37	1.35	1.18	1.61	1.47	1.49	1.44	1.28	1.14	0.71
	SE	1.37	1.35	1.18	1.61	1.47	1.49	1.44	1.28	1.14	0.71
	Cohen's d	0.43	0.39	0.38	0.19	0.37	0.43	0.32	0.39	0.28	0.11
	Vortioxetine	-1.82	-2.12	-1.30	-1.60	-1.05	-1.63	-1.67	-1.58	-1.38	-0.60
	20 mg (n = 73)	1.29	1.41	1.39	1.48	1.55	1.55	1.42	1.35	1.25	0.76
	SE	1.29	1.41	1.39	1.48	1.55	1.55	1.42	1.35	1.25	0.76
Cohen's d	0.50	0.49	0.27	0.21	0.28	0.37	0.41	0.51	0.38	0.24	
Patients with baseline QIDS-J ≤ 10 /HSAS ≥ 20	Placebo (n = 98)	-1.57	-1.83	-1.23	-1.32	-1.04	-1.73	-1.63	-1.53	-1.34	-0.54
	Change										
	SE	1.25	1.50	1.27	1.54	1.37	1.34	1.33	1.45	1.39	0.92
	Vortioxetine	-1.78	-1.90	-1.37	-1.56	-1.51	-1.62	-1.50	-1.59	-1.36	-0.53
	10 mg (n = 98)	1.30	1.58	1.44	1.58	1.35	1.34	1.36	1.34	1.28	0.84
	SE	1.30	1.58	1.44	1.58	1.35	1.34	1.36	1.34	1.28	0.84
	Cohen's d	0.16	0.05	0.10	0.16	0.35	-0.08	-0.10	0.04	0.02	-0.01
	Vortioxetine	-1.90	-2.08	-1.79	-1.32	-1.31	-1.85	-1.66	-1.57	-1.30	-0.69
	20 mg (n = 87)	1.41	1.53	1.22	1.37	1.62	1.60	1.36	1.49	1.24	0.75
	SE	1.41	1.53	1.22	1.37	1.62	1.60	1.36	1.49	1.24	0.75
Cohen's d	0.24	0.17	0.45	0.00	0.18	0.08	0.02	0.03	-0.03	0.18	

Abbreviations: MADRS, Montgomery-Åsberg Depression Rating Scale; rQIDS-J, revised Japanese version of the Quick Inventory of Depressive Symptomatology incorporating the Himorogi Self-rating Anxiety Scale (HSAS); SE, standard error.

Note: MADRS01, reported sadness; MADRS02, apparent sadness; MADRS03, inner tension; MADRS04, reduced sleep; MADRS05, reduced appetite; MADRS06, concentration difficulties; MADRS07, lassitude; MADRS08, inability to feel; MADRS09, pessimistic thoughts; MADRS10, suicidal thoughts.

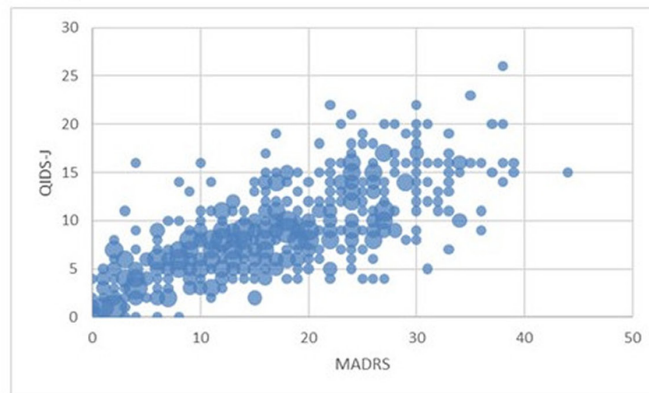


Visit 3 (Baseline)



$$R^2 = 0.17$$

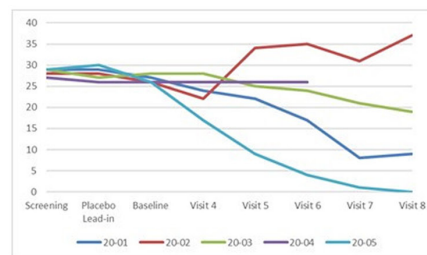
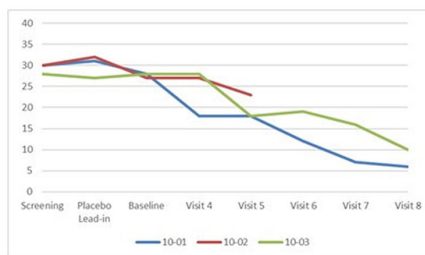
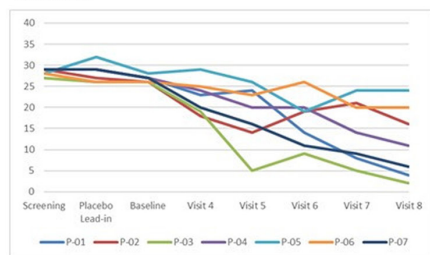
Visit 8



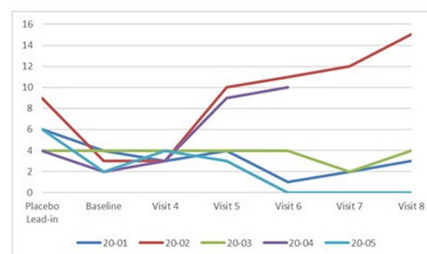
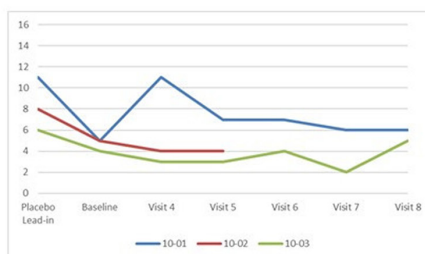
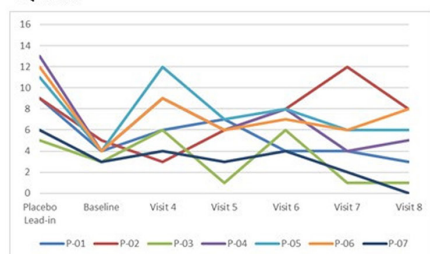
$$R^2 = 0.45$$

FIGURE 2 Correlation between MADRS and QIDS-J scores at baseline (visit 3) and at week 8 (visit 8)

MADRS



QIDS-J

FIGURE 3 Change from baseline in MADRS total scores among patients with baseline QIDS-J scores ≤ 5 in the CCT-004 trial, suggesting slight baseline score inflation

in QIDS-J (thus suggesting no inconsistency within QIDS-J), and “reduced appetite” was shown to improve in MADRS as well, thus suggesting no dissociation between QIDS-J and MADRS. Additionally, study results also suggested that the smaller the appetite improvement in patients, the more likely they are to exhibit a low placebo response, while the greater their appetite improvement, the more likely they are to exhibit a high placebo response (Figure S1).

Research also suggests that there is a tendency among trial investigators to overestimate the severity of depression at baseline in their determination to achieve their accrual target, currently known as “baseline score inflation,” which is thought likely to increase the placebo response and decrease the probability of signal detection (i.e., detection of the therapeutic potential of any antidepressant being tested).⁶

In the CCT-004 trial, an examination of the MADRS and QIDS-J scores revealed that the MADRS total scores and the QIDS-J scores were

more positively correlated at week 8 (visit 8) ($r = 0.45$) than at baseline (visit 3) ($r = 0.17$) (Figure 2), while an examination of patients with baseline QID-J scores ≤ 5 (i.e., those likely without depression) in all groups showed that the MADRS and QIDS-J ratings differed in many of these patients not only between screening and baseline but between baseline and visit 4 (Figure 3), suggesting baseline score inflation, albeit minor.

The current controversy surrounding placebo response in clinical trials of MDD may be summarized as an ongoing debate on whether the treatment and placebo responses are correlated. If they are correlated, a high placebo response is associated with a high treatment response. On the other hand, if the placebo response varies between the placebo and treatment arms depending on baseline depression severity, and when higher/lower in the placebo arm, leads to the observed treatment effect being lower/higher.^{3,5,17}

Indeed, according to an analysis by Khan et al¹⁸ of the US Food and Drug Administration (FDA) data for 52 placebo-controlled



randomized clinical trials (RCTs) of antidepressants, only 21.1% of the trials reporting a high placebo response demonstrated statistically significant symptomatic improvements with antidepressants over placebo, while as many as 74.2% of the trials reporting a low placebo response demonstrated statistically significant symptomatic improvements with antidepressants over placebo. It is thus generally suggested that the greater the placebo response, the more difficult it becomes for active drugs to show statistically significant symptomatic improvements versus placebo.

Now, mirtazapine may be of interest here as an antidepressant with a unique adverse event profile (i.e., somnolence and excessive sedation),^{19,20} which might allow its action to be detected even in double-blind trials, likely tipping the MADRS or HAMD 17 ratings in favor of mirtazapine over placebo and thus accounting for a smaller placebo response than other antidepressants. Despite this assumption, however, mirtazapine did fail to deliver statistically significant symptomatic improvements over placebo in a 6-week, placebo-controlled trial in hospitalized patients with MDD,²¹ with the change in HAMD 17 total score from baseline (primary endpoint) shown to be 15.1 and 12.5 with mirtazapine and placebo, respectively. According to the investigators, this failure was attributable to a number of methodological flaws inherent in the study, which included: the up-titration of mirtazapine over as short a period as 6 weeks, which led to the drug being given at an insufficient dose for an insufficient duration thus resulting in as many dropouts from among those given mirtazapine due to insufficient efficacy as from among those given placebo; the limited number of patients registered at some of the study sites; and the unavailability of concomitant benzodiazepines for use from week 3 onwards.²¹ However, a comparison of this failed study,²¹ a review of 5 mirtazapine trials by Davis et al,²² and a review of mirtazapine trial meta-analyses by Kasper²³ shows that while the change in HAMD 17 score with mirtazapine in the failed study did not differ greatly from that in the other mirtazapine trials,^{22,23} the change in HAMD 17 score with placebo in the failed study was greater than that in the other mirtazapine trials,^{22,23} suggesting again that the failure of mirtazapine to deliver significant symptomatic improvements could be accounted for not so much by the methodological flaws involved as by the large placebo response shown in those treated with placebo.

Thus, while it remains unclear whether the placebo effect may vary depending on the baseline depression severity or if there may be an increase in treatment effect for patients with more severe depression, the fact remains that placebo response represents an overarching issue to be resolved in clinical trials of MDD, providing a rationale for further exploring the optimal use of clinical measures and methods of analysis.

In this regard, while recent efforts focused on designing trials to reduce large placebo responses or remove high placebo responders may be found ineffective,⁵ our study may represent a viable approach in this line of research.

Earlier studies in the literature suggest that factors contributing to placebo responses in clinical trials of MDD include number of study sites, number of patients per study site, number of depressive episodes per patient, and severity of depression in study subjects at

enrollment; and that anxiety disorder of varying severity co-exists in a certain proportion of subjects with MDD thus contributing to placebo responses in clinical trials of MDD.

Again, while the placebo run-in design should be useful in excluding subjects who exhibit symptomatic improvement or deterioration early in their natural course during single-blind placebo treatment, there are other factors contributing to the placebo response among the subjects. These include investigator-induced baseline score inflation, which may remain in place through the single-blind placebo treatment phase; recruitment of subjects from a heterogeneous population of patients with MDD, which may lead to those with quasi-bipolar disease or quasi-MDD being included. Thus, to avoid this from occurring, the CCT-004 trial included those with recurrent MDD only to the exclusion of those with refractory MDD by qualifying the duration of depression symptoms at study enrollment to ensure a high probability of success of the study.

In this light, our study may be of particular interest, in that it was designed to evaluate patients with depression for severity of not only MDD but also anxiety disorder at baseline and over time, using both self-administered and objective ratings as mediated by central monitoring to optimize patient selection and monitoring thereby minimizing placebo responses in the trial.

It is suggested that the placebo response may vary depending on the severity of anxiety disorder as a comorbidity in patients with depression enrolled in depression clinical trials,²⁴ with this response shown to be the smallest in patients with obsessive-compulsive disorder (OCD), of all patients with anxiety disorder,²⁵ and that social anxiety disorder as a comorbidity contributes to a lower placebo response in patients with MDD.²⁶

In this regard, the HSAS¹³ used in this study was developed to quantify the severity of anxiety to evaluate its severity in reference to the Hamilton Anxiety Rating Scale (HAM-A). Given that, generally, the more severe their panic disorder or generalized anxiety disorder (GAD), the higher their HSAS scores in patients with MDD, it is likely that high HSAS scores among these patients contributed to a higher placebo response in this study.

Study results showed that the placebo responses, as well as the HSAS cutoffs, varied depending not only on the severity or kind of anxiety disorder present in some patients with MDD but also on the properties/profile of the antidepressant evaluated. Thus, the QIDS-J/HSAS cutoffs identified in this study as contributing or not contributing to the placebo response may not apply in other clinical trials evaluating different antidepressants. The effect size may widely differ between the active treatment (e.g., venlafaxine) and placebo in patients in whom anxiety disorder proved a major component²⁷ or may not differ between the active treatment (e.g., duloxetine) and placebo, in all patients evaluated, regardless of severity of their anxiety disorders.²⁶

Findings from this study suggest that it may be worthwhile to collect and analyze HSAS data from phase 2 trials in patients with MDD evaluated for severity of depression and anxiety with QIDS-J and HSAS to help identify appropriate combinations of QIDS-J/HSAS scores (severity of depression/anxiety) for candidates for



phase 3 studies, thus allowing inclusion of patients whose severity of depression (moderate to severe depression) and anxiety would bring out the best in any antidepressant being evaluated, as well as identification of patient profiles that would contribute to success in clinical trials with any particular antidepressant or those to be targeted for treatment with such antidepressant after its approval.

Finally, overall, current study findings appear to point to the importance of designing clinical trials of MDD, with a focus on the need to evaluate patients with MDD for presence and severity of coexisting anxiety disorder at study enrollment.

Furthermore, recent research also suggests a role for centralized rating, as opposed to site-based rating, in facilitating selection of candidates for clinical trials of MDD, as well as in reducing placebo responses.⁶ Likewise, the use of "central monitoring" in the CCT-004 trial not only helped reduce the baseline score inflation by calling the study sites' attention to their subjects who showed a large discrepancy in their clinical course as assessed by QIDS-J versus MADRS, but also led to a number of interesting findings, including the observation that those with disproportionately low baseline QIDS-J scores were associated with a large placebo response. It is of note here that the investigators in the CCT-004 trial were properly qualified in their skills, that is, included only if they were found to have had prior experience with depression clinical trials and MADRS evaluations and that the study had in place an MADRS accreditation scheme in which they participated in hands-on sessions on MADRS evaluations using a uniform interview video to ensure standardization and calibration of their skills as clinical examiners prior to study initiation.

There are some study limitations to be considered. First, this was a post hoc exploratory analysis of the data from the CCT-002, CCT-003, and CCT-004 trials of the antidepressant vortioxetine. Second, while central monitoring in this study identified and helped minimize the discrepancy between self-administered and objective ratings, it remains unclear how to define or what to make of any discrepancy between these ratings and further research is required to elucidate the implications of such discrepancy for clinical research and practice. Thus, the study findings reported here need to be interpreted with caution and call for further study in randomized, placebo-controlled trials.

Despite these limitations, however, this study appears to provide several key findings that could open a new avenue for patient selection and monitoring in depression clinical trials, where placebo responses tend to mask the therapeutic potential of antidepressants.

5 | CONCLUSION

the use of both physician and patient ratings of depression and anxiety in MDD patients at baseline and over time and their central monitoring, as well as a close examination of their subscales and their clinical course, was shown to play a role in minimizing baseline score

inflation and optimizing patient selection at baseline and patient monitoring over time in clinical trials thus allowing antidepressants to be evaluated for their full therapeutic potential. Again, Japanese MDD patients with baseline QIDS-J scores ≥ 11 /HSAS ≤ 19 may be targeted for inclusion in depression clinical trials.

AUTHOR CONTRIBUTIONS

YW, AN, TK, NS, MI, II, and KW conceptualized and designed the study and revised the manuscript. AN, MI, and II contributed to data acquisition, analyzed the data and drafted the figures. YW, AN, MI, and II drafted the manuscript.

ACKNOWLEDGMENTS

Yoshinori Watanabe passed away shortly after having reviewed and approved the manuscript for publication. Medical writing/editorial assistance was provided by Hiroaki Itoh of Interface, Kanagawa, Japan and funded by Takeda Pharmaceutical Co., Ltd.

FUNDING INFORMATION

This study was supported by Takeda Pharmaceutical Co., Ltd.

CONFLICT OF INTEREST

In the CCT-004 trial, Yoshinori WATANABE, Koichiro WATANABE, Toshiaki KIKUCHI, and Norifusa SAWADA provided counsel on efficacy assessments and served as medical officers responsible for central monitoring and Yoshinori WATANABE declared no other conflict of interest. Akira NISHIMURA is a former employee of Takeda Pharmaceutical Co., Ltd, and Manami IMAZAKI and Isao INAD are current employees of Takeda Pharmaceutical Co., Ltd, and they declare no other conflicts of interest. Koichiro WATANABE has received grant funding and consultant fees from Otsuka Pharmaceutical Co., Ltd., Pfizer Japan Inc., Shionogi & Co., Ltd., Sumitomo Pharma (formerly Sumitomo Dainippon Pharma), Taisho Pharmaceutical Co., Ltd. (Taisho Toyama Pharmaceutical Co., Ltd.), and Takeda Pharmaceutical Co., Ltd. Norifusa SAWADA has received speaker honoraria from Janssen Pharmaceutical Co., Ltd., Sumitomo Pharma (formerly Sumitomo Dainippon Pharma) as well as consultant fees from Takeda Pharmaceutical Co., Ltd. Toshiaki KIKUCHI has received consultant fees from Lundbeck, Sumitomo Pharma (formerly Sumitomo Dainippon Pharma), Eli Lilly, Kyowa Kirin Co., Ltd. (formerly Kyowa-Hakko Kirin Co., Ltd.), Meiji Seika Pharma Co., Ltd., Mochida Pharmaceutical Co., Ltd., MSD K.K., and Otsuka Pharmaceutical Co., Ltd.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from Medical Affairs, Takeda Pharmaceutical Co., Ltd. but restrictions apply to the availability of these data, as they were used under license for this study, and so are not publicly available. However, the data are made available through Vivli at the Takeda website <https://clinicaltrials.takeda.com/takedas-commitment?commitment=5> upon reasonable request from bona fide qualified researchers and with the permission of Takeda Pharmaceutical Co., Ltd. "cd_value_code="text.

ETHICS APPROVAL

The IRB of the CCT-004 trial approved the protocol of this post hoc analysis as well as the request to waive informed consent for this analysis as involving no more than minimal risk to the participants in the trial.

ORCID

Akira Nishimura  <https://orcid.org/0000-0003-2064-8421>

REFERENCES

- Mancini M, Wade AG, Perugi G, Lenox-Smith A, Schacht A. Impact of patient selection and study characteristics on signal detection in placebo-controlled trials with antidepressants. *J Psychiatr Res*. 2014;51:21–9.
- Khin NA, Chen YF, Yang Y, Yang P, Laughren TP. Exploratory analyses of efficacy data from major depressive disorder trials submitted to the US Food and Drug Administration in support of new drug applications. *J Clin Psychiatry*. 2011;72:464–72.
- Khan A, Leventhal RM, Khan SR, Brown WA. Severity of depression and response to antidepressants and placebo: an analysis of the food and drug administration database. *J Clin Psychopharmacol*. 2002;22(2):40–5.
- Marchesi C, De Panfilis C, Tonna M, Ossola P. Is placebo useful in the treatment of major depression in clinical practice? *Neuropsychiatr Dis Treat*. 2013;9:915–20.
- Whitlock ME, Woodward PW, Alexander RC. Is high placebo response really a problem in depression trials? A critical re-analysis of depression studies. *Innov Clin Neurosci*. 2019;16:12–7.
- Kobak KA, Leuchter A, DeBrotta D, Engelhardt N, Williams JBW, Cook IA, et al. Site versus centralized raters in a clinical depression trial: impact on patient selection and placebo response. *J Clin Psychopharmacol*. 2010;30:193–7.
- Kendrick T, El-Gohary M, Stuart B, et al. Routine use of patient reported outcome measures (PROMs) for improving treatment of common mental health disorders in adults. *Cochrane Database Syst Rev*. 2016;7:CD011119.
- Herschenberg R, McDonald WM, Crowell A, et al. Concordance between clinician-rated and patient reported outcome measures of depressive symptoms in treatment resistant depression. *J Affect Disord*. 2020;266:22–9.
- Doraiswamy PM, Bernstein IH, Rush AJ, Kyutoku Y, Carmody TJ, Macleod L, et al. Diagnostic utility of the Quick Inventory of Depressive Symptomatology (QIDS-C16 and QIDS-SR16) in the elderly. *Acta Psychiatr Scand*. 2010;122:226–34.
- Elmquist JM, Melton TK, Croarkin P, McClintock SM. A systematic overview of measurement-based care in the treatment of childhood and adolescent depression. *J Psychiatr Pract*. 2010;16:217–34.
- Inoue T, Sakai K, Kitagawa T, Nishimura A, Inada I. Randomized, double-blind, placebo-controlled study to assess the efficacy and safety of vortioxetine in Japanese patient with major depressive disorder. *Psychiatry Clin Neurosci*. 2020;74:140–8.
- Rush AJ, Bernstein IH, Trivedi MH, Carmody TJ, Wisniewski S, Mundt JC, et al. An evaluation of the quick inventory of depressive symptomatology and the hamilton rating scale for depression: a sequenced treatment alternatives to relieve depression trial report. *Biol Psychiatry*. 2006;59:493–501.
- Mimura C, Nishioka M, Sato N, Hasegawa R, Horikoshi R, Watanabe Y. A Japanese scale to assess anxiety severity: development and psychometric evaluation. *Int J Psychiatry Med*. 2011;41:29–45.
- Piercy MA, Sramek JJ, Kurtz NM, Cutler NR. Placebo response in anxiety disorders. *Ann Pharmacother*. 1996;30:1013–9.
- Peciña M, Bohnert AS, Sikora M, et al. Association between placebo-activated neural systems and antidepressant responses: neurochemistry of placebo effects in major depression. *JAMA Psychiatr*. 2015;72:1087–94.
- Rush AJ, Carmody TJ, Reimnitz PE. The inventory of depressive symptomatology (IDS): clinician (IDS-C) and self-report (IDS-SR) ratings of depressive symptoms. *Int J Methods Psychiatr Res*. 2000;9:45–59.
- Fournier JC, DeRubeis RJ, Hollon SD, et al. Antidepressants drug effects and depression severity: a patient-level meta-analysis. *JAMA*. 2010;303:47–53.
- Khan A, Detke M, Khan SRF, Mallinckrodt C. Placebo response and antidepressant clinical trial outcome. *J Nerv Ment Dis*. 2003;191:211–8.
- Fawcett J, Barkin RL. Review of the results from clinical studies on the efficacy, safety and tolerability of mirtazapine for the treatment of patients with major depression. *J Affect Disord*. 1998;51:267–85.
- Watanabe N, Omori IM, Nakagawa A, Cipriani A, Barbui C, Churchill R, et al. Mirtazapine versus other antidepressive agents for depression. *Cochrane Database Syst Rev*. 2011;12:CD006528.
- Vartiainen H, Leinonen E. Double-blind study of mirtazapine and placebo in hospitalized patients with major depression. *Eur Neuropsychopharmacol*. 1994;4:145–50.
- Davis R, Wilde MI. Mirtazapine: a review of its pharmacology and therapeutic potential in the management of major depression. *CNS Drugs*. 1996;5:389–402.
- Kasper S. Clinical efficacy of mirtazapine: a review of meta-analyses of pooled data. *Int Clin Psychopharmacol*. 1995;10(Suppl 4):25–35.
- Huppert JD, Schultz LT, Foa EB, Barlow DH, Davidson JRT, Gorman JM, et al. Differential response to placebo among patients with phobia, panic disorder, and obsessive-compulsive disorder. *Am J Psychiatry*. 2004;161:1485–7.
- Lyndon GJ, Prieto R, Wajsbrot DB, Allgulander C, Bandelow B. Efficacy of venlafaxine extended release in major depressive disorder patients: effect of baseline anxiety symptom severity. *Int Clin Psychopharmacol*. 2019;34:110–8.
- Nelson JC. Anxiety does not predict response to duloxetine in major depression: results of a pooled analysis of individual patient data from 11 placebo-controlled trials. *Depress Anxiety*. 2010;27:12–8.
- Strawn JR, Dobson ET, Mills JA, Cornwall GJ, Sakolsky D, Birmaher B, et al. Placebo response in pediatric anxiety disorders: results from the child/adolescent anxiety multimodal study. *J Child Adolesc Psychopharmacol*. 2017;27:501–8.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Watanabe Y, Nishimura A, Kikuchi T, Sawada N, Imazaki M & Inada I et al. Central monitoring of depression and anxiety symptoms reduces placebo responses in depression clinical trials: A post hoc exploratory analysis of data from the phase III CCT-004 trial of vortioxetine. *Neuropsychopharmacol Rep*. 2022;42:468–477. <https://doi.org/10.1002/npr2.12288>