Development and validation of a prediction model for the probability of responding to placebo in antidepressant trials: a pooled analysis of individual patient data

Kiyomi Shinohara,[1] Shiro Tanaka,[2] Hissei Imai,[1] Hisashi Noma,[3] Kazushi Maruo,[4] Andrea Cipriani,[5,6] Shigeto Yamawaki,[7] Toshi A Furukawa[1]

ABSTRACT

Background Identifying potential placebo responders among apparent drug responders is critical to dissect drug-specific and nonspecific effects in depression.
Objective This project aimed to develop and test a prediction model for the probability of responding to placebo in antidepressant trials. Such a model will allow us to estimate the probability of placebo response among drug responders in antidepressants trials.
Methods We identified all placebo-controlled, double-blind randomised controlled trials (RCTs) of second generation antidepressants for major depressive disorder conducted in Japan and requested their individual patient data (IPD) to pharmaceutical companies. We obtained IPD (n=1493) from four phase II/III RCTs comparing mirtazapine, escitalopram, duloxetine, paroxetine and placebo. Out of 1493 participants in the four clinical trials, 440 participants allocated to placebo were included in the analyses. Our primary outcome was response, defined as 50% or greater reduction on Hamilton Rating Scale for Depression at study endpoint. We used multivariable logistic regression to develop a prediction model. All available candidate of predictor variables were tested through a backward variable selection and covariates were selected for the prediction model. The performance of the model was assessed by using Hosmer-Lemeshow test for calibration and the area under the ROC curve for discrimination.
Findings Placebo response rates differed between 31% and 59% (grand average: 43%) among four trials. Four variables were selected from all candidate variables and included in the final model: age at onset, age at baseline, bodily symptoms, and study-level difference. The final model performed satisfactorily in terms of calibration (Hosmer-Lemeshow p=0.92) and discrimination (the area under the ROC curve (AUC): 0.70).
Conclusions Our model is expected to help researchers discriminate individuals who are more likely to respond to placebo from those who are less likely so.
Clinical implications A larger sample and more precise individual participant information should be collected for better performance. Examination of external validity in independent datasets is warranted.
Trial registration number CRD42017055912.

INTRODUCTION

Placebo response has an important role in antidepressant trials.[1] In the past 15 years, there has been a long-lasting discussion about whether placebo response has been increasing over time.[2] This could have contributed to the increasing rates of failures of antidepressants trials.[3] However, recent and more sophisticated analyses revealed that placebo response has actually been stable since 1990s[1] and that response to placebo is influenced by many factors such as baseline severity, the length of study and the proportion of participant allocated to placebo.[4] Thus, careful attention to relevant factors is necessary for examining placebo response.

Estimating placebo response is also critical to dissect drug-specific and non-specific effects in neurobiological studies. In a clinical trial, response to drug can occur in two types of patients.[5 6] The first type represents patients who responded to the drug but would not have responded to the placebo: they are 'drug only responders'.[6] The second type is those who would have responded both on drug and placebo: they are 'always responders 'and these are 'placebo responders'. If we study biological changes among the apparent drug responders, we would be mixing these two types of patients, and the effect of the drug would be diluted. However, as placebo and drugs cause overlapping changes,[7] the ability to distinguish drug-specific effect from placebo effect would still be limited.[8] If we can identify 'placebo responders' among apparent drug responders, researchers will be able to focus on 'drug only responders' and assessment of the true drug effect will be more straightforward and accurate.[6]

Previous reviews showed that both design-related factors (eg, multicentre and study duration) and patient factors (baseline severity, sex and age) relate to placebo response.[4 9–12] However, they identified these factors based on the aggregate data meta-analysis of clinical trials. That is, these factors were derived from the difference of average on outcomes, and such analyses are subject to ecological fallacy as the relationships observed at study level may not be applicable to the individual level.[13] We need individual participant data (IPD) to investigate patient-level effect moderators. To the best of our knowledge, there are only two studies to date that aimed at predicting placebo response in depression trials by using individual-level dataset.[14 15] Both studies suggested that younger age and less severe

depression were associated with high placebo response. Their models showed moderate performance, as indicated by the area under the receiver operating characteristic (ROC) curve (AUC) of around 0.6 in internal validation test. However, some important factors, such as age at onset of the first episode and physical illness, were not fully examined in their dataset. Also, their dataset included one trial only[14] or duloxetine trials only.[15] Thus, their models have limited value in terms of validity and generalisability.

Our objective is to develop and test a model to predict the probability of placebo response retrospectively among apparent drug responders in clinical trials. Although we cannot predict placebo response prospectively, estimating the probability of placebo response of individual drug responder with this model will enable researchers to distinguish 'drug only responders' from 'placebo responders' among apparent drug responders. Our model is expected to help identify biomarkers of drug response in biological studies.

EXPERIMENTAL PROCEDURES

We followed the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis statement[16] for reporting development and validation of this model. As we used the anonymised individual data from clinical trials, this study did not need approval from ethics boards. The present study is one of the projects in IPD meta-analysis of antidepressant trials for major depression in Japan (AD-Japan) and had been defined in the protocol of AD-Japan before commencement.[17]

Data source

We obtained IPD from pharmaceutical companies through the public–private partnerships (PPP) with the International College of Neuropsychopharmacology and the Japanese Society of Neuropsychopharmacology after complete anonymisation. In AD-Japan, we included all placebo-controlled, double-blind randomised trials of new generation antidepressants in the acute phase treatment of major depression conducted in Japan. We focused on the trial in Japan in order to maximise the likelihood of obtaining IPD of all related trials. We searched eligible trials in the recently published systematic review and a network meta-analysis of 21 antidepressants.[18] We identified 11 trials by six pharmaceutical companies and requested their IPDs through PPP. Four companies agreed to provide the data, and three have provided the data of four trials that enabled us to analyse directly: Meiji Seika Pharma (one study, n=270), Mochida Pharmaceutical (two studies, n=298 and n=485) and Shionogi (one study, n=440).[17] We used the data from these four randomised controlled trials (RCTs) comparing mirtazapine, escitalopram, duloxetine with placebo or paroxetine in the treatment for depression[19–22] (table 1). Among 1493 individuals participating in the four clinical trials, 440 were allocated to placebo. Trials were conducted in multiple sites (45–84 sites per study) in Japan between 2004 and 2010, and the trial duration was 6–8 weeks. Table 1 lists the design of the four trials. All trials used Hamilton Rating Scale for Depression (HRSD)[23] for both baseline and endpoint measurements. Two trials used placebo run-in.[19 21]

Participants

All participants who were included in the full analysis set and allocated to placebo arm were included in the current analyses. The primary eligibility criteria of participants were very similar across different studies: age range between 20 years and 65 years (75 years in the mirtazapine study,[20]) with the diagnosis of major depression according to DSM-IV or DSM-IV-TR.

Table 1 Characteristics of included studies

	Duloxetine P3 trial	Mirtazapine P3 trial	Escitalopram P2 trial	Escitalopram P3 trial	All trials
Dosage (per day) and controls	Duloxetine 40 mg. Duloxetine 60 mg. Paroxetine 20–40 mg. Placebo.	Mirtazapine 15 mg. Mirtazapine 30 mg. Mirtazapine 45 mg. Placebo.	Escitalopram 10 mg. Escitalopram 20 mg. Placebo.	Escitalopram 10 mg. Escitalopram 20 mg. Paroxetine 40 mg. Placebo.	
Minimum severity at baseline	HRSD-17 ≥19	HRSD-17 ≥18	HRSD-17 ≥18	MADRS ≥22	
Number of sites	84	45	47	59	62.3 (SD 16.1)
Number of arms	4	4	3	4	4* (SD 0.4)
Placebo run-in (%)	100†	0	0	100‡	61.1
Fixed dose (%)	0§	100	100	100	67.0
Length of trial (weeks)	6	6	8	8	7 (SD 1)
Sample size	440	270	298	485	393.0 (SD 87.2)
Placebo sample size	145	70	101	124	110 (SD 27.9)
Proportion of participants allocated to placebo (%)	33.0	25.9	33.9	25.6	30.0
Response rate in placebo arm (%)	37.9	50.0	59.4	31.5	43.0

*Median.
†In duloxetine P3 trial, placebo run-in was after inclusion. There were no criteria of exclusion at the end of placebo run-in phase.
‡In escitalopram P3 study, patients with reduction of MADRS >25% or CGI-S ≤3 were excluded.
§As paroxetine dosage was flexible, dosage appeared to be flexible in all study arms even though duloxetine was administered as a fixed dose.
HRSD, Hamilton Rating Scale for Depression; MADRS, Montgomery-Asberg Depression Rating Scale.
P2 : Phase 2 and P3: Phase 3; CGP - Clinical Global Impression – Severity scale

Outcome

As we aimed to identify placebo responders among apparent drug responders in biological studies, we decided our primary outcome was response. Response was defined as 50% or greater reduction on HRSD between baseline and study endpoint. All the trials were double blind, and the assessor was masked.

Candidates of predictor variables

Previous studies have suggested various factors contributing to response to placebo.[9–11] We selected the following variables based on the literature review[9–11 24] and the availability of data in the current dataset: sex*, inpatient status*, age at onset, time since first onset (time between first episode and age at baseline), age at baseline, number of previous episodes (1/2/3 or more)*, duration of current episode, HRSD at baseline, features of a current episode such as melancholic*, stress-precipitated (details were described below), comorbidity of physical illness*, prior use of antidepressants* and the use of rescue medication during the trial* (*categorical variables. Others are numerical variables). We also assessed the following prespecified symptomatological subscales derived from the previous factor analysis of 17-item HRSD to test the association between depressive symptoms and placebo response[25]: insomnia (items 4, 5 and 6), anhedonia/retardation (items 1, 7, 8 and 14) and bodily symptoms (items 11, 13 and 15). Factors related to recent stress were measured only in two trials, and they classified the type of stress into nine categories, which we converted into a dichotomous variable (any stress/no stress)* and a continuous score (the sum of items). No study assessed trauma or other past stressful event.

Model

We used multivariable logistic regression to assess the probability of response to placebo. As we included only four trials, we used fixed effects for adjusting for study-level baseline difference. Candidates of covariates in the model were selected through a backward variable selection with the critical value of p=0.15. We developed the final model (model 1) with those selected variables. In addition to the model 1, we developed supplementary models by removing a variable and keeping the most important and necessary variables to evaluate the influence of specific variables. We used the completers' dataset for the main analysis because we aimed to make a model for trials using biological measurements and they usually target individuals who have completed the assessments.

Model performance

The performance of the model was assessed based on the probability of response on placebo. Hosmer-Lemeshow test was used as a measure for calibration, and AUC was used as a measure of discrimination. We evaluated internal validity of the final multivariable logistic regression with the use of leave-one-trial-out cross-validation. We used SAS V.9.4 for statistical analysis.

Sensitivity analysis

We developed a model after multiple imputation (MI) of missing values in order to test the robustness of the results based on the completers. First, we created 100 multiple copies of the dataset with the missing values replaced by imputed value.[26] Then we performed logistic regression with the selected variables in the final model for each dataset and pooled the results.

RESULTS

Characteristics of the participants

Figure 1 shows the flow chart of participants, and table 2 summarises baseline characteristics of the participants. The average age was 37.4, and 47.5% was women. The average baseline severity on 17-item HRSD was 21.1, and 39% of participants had a history of past depressive episodes. The primary outcome was missing in 63 participants (14% of those allocated to placebo) at the endpoint. Among individuals who completed the assessment at the endpoint (n=377), 189 participants responded. Placebo response rates differed considerably among trials, ranging between 32% and 59% with an average rate of 43% (table 1).

Model and selected variables

The logistic regression identified four variables through backward variable selection (table 3): age at onset, age at baseline, bodily symptoms and study (univariate models as preliminary analysis are in online supplementary table S1). The OR of 'age at onset' and 'age at baseline' for placebo response were 0.66 (per 10 years) and 1.97 (per 10 years), respectively. To explore the association of age-related factors, we inserted age at baseline (model 2), age at onset (model 3) and time since first onset (model 4) separately into the final model based on the same completers' dataset. The results showed that elderly participants are less likely to respond to placebo given the same age of onset because they had longer time since onset. Whereas if they have the same period after the first episode, the older group responded better than the younger group. It means that age at baseline has
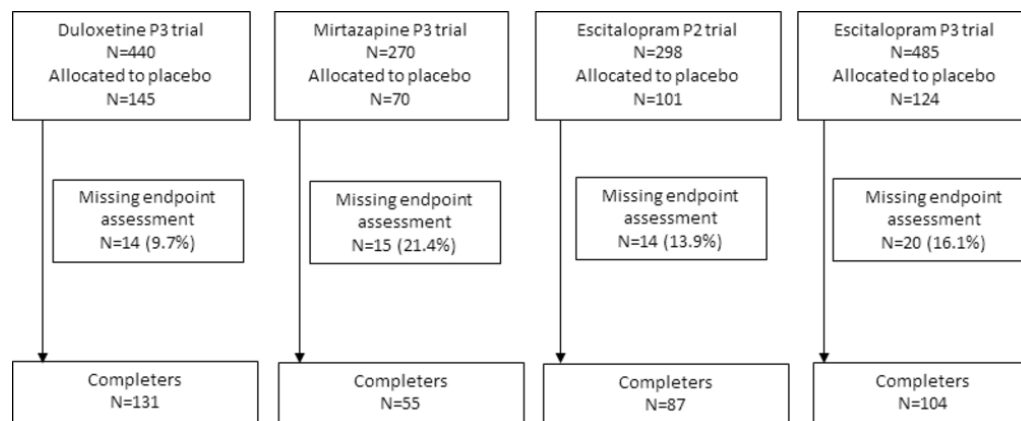

Figure 1   Study cohort from four antidepressants trials.

Table 2 Characteristics of patients included in the analyses (placebo arms)

	Total (n=440)			Duloxetine P3 trial (n=145)		Mirtazapine P3 trial (n=70)		Escitalopram P2 trial (n=101)		Escitalopram P3 trial (n=124)	
	Mean	SD	Missing	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Female (%)	47.5		0	45.5		44.3		47.5		51.6	
Inpatient (%)	0.9		0	0.7		4.3		0.0		0.0	
Age at onset (years)	34.0	11.4	2	35.3	11.3	36.3	13.1	32.4	9.5	32.3	11.7
Age at baseline (years)	37.4	10.8	0	38.7	10.5	39.9	12.8	35.0	8.9	36.4	10.7
Time since first onset (years)	3.5	5.0	2	3.4	5.2	3.9	5.8	2.6	3.7	4.0	5.2
No. of previous episodes:											
1 (%)	61.1		6	61.4		70.3		59.4		57.3	
2 (%)	22.8		6	22.1		20.3		27.7		21.0	
3 or more (%)	16.1		6	16.6		9.4		12.9		21.8	
Duration of current episode (weeks)	44.9	66.3	6	44.7	21.4	54.7	77.5	24.6	46.9	56.4	97.6
HRSD (17 items) at baseline	21.1	4.1	0	20.4	4.2	22.5	3.6	22.5	3.6	20.0	4.2
Melancholic (%)	84.6		70	86.9	–		–	85.1	–	81.5	–
Anhedonia/retardation	7.6	1.7	0	7.4	1.8	8.0	1.8	8.1	1.4	7.4	1.8
Bodily symptoms	3.8	1.5	0	3.9	1.4	3.9	1.3	3.9	1.4	3.4	1.6
Insomnia	3.0	1.7	0	2.7	1.7	3.3	1.8	3.1	1.7	3.2	1.6
Any stress (%)	76.0		215	–		–		86.1		67.7	
Score of stress (0–9)	1.1	0.9	215	–		–		1.2	0.8	1.0	1.0
Comorbidity of physical illness (%)	61.6		0	67.6		48.6		50.5		71.0	
Prior use of antidepressants (%)	31.5		145	—		18.6		17.8		50.0	
Number of antidepressants	0.7	1.1	202	–		2.1	0.9	0.3	0.6	0.9	1.2
Use of rescue medication (%)	56.4		0	47.6		35.7		72.3		65.3	

HRSD, Hamilton Rating Scale for Depression.

a positive influence on placebo response, but the length of the period between the first episode and current episode is more important than the age. The older individuals who had first developed depression in their youth were least likely to respond to placebo. Regarding the severity of depression, instead of HRSD total score, the bodily symptoms, a subscale score from HRSD, was included in the final model. When we inserted HRSD total score instead of bodily symptoms, the coefficient was almost the same (OR=0.95) (model 5 in online supplementary table S2). Among the four factors, the difference in trials showed the strongest influence on the placebo response rate. If we take the

mirtazapine trial as the comparator, the OR of placebo response ranged between 0.8 and 3.3 in the other trials.

The final model for calculating the probability of placebo response, p, is expressed as the equation below: $P=1/(1+\exp(-0.4853+0.0412*x1-0.0677*x2+0.1656*x3+0.9925*x4-0.201*x5+1.1702*x6))$, where x1 is age at baseline, x2 is age at onset, x3 is the bodily symptoms at baseline, x4–x6 are dummy variables for trials (x4=1 if duloxetine P3 trial, x5=1 if escitalopram P2 trial, x6=1 if escitalopram P3 trial, and these variables are set at 0 otherwise). In the sensitivity analysis using MI, the same four variables were selected.

Table 3 Multivariable logistic regression analysis of placebo response and predictors

	Model 1: final model (n=375*)			Model 2: adjusted for age at baseline (n=377)			Model 3: adjusted for age at onset (n=375)			Model 4: adjusted for time since onset (n=375)		
	OR	95% CI	P value	OR	95% CI	P value	OR	95% CI	P value	OR	95% CI	P value
Age at baseline (for every 10 years)	0.66	0.40 1.10	0.111	1.28	1.04 1.58	0.020	–			–		
Age at onset (for every 10 years)	1.97	1.22 3.17	<0.01	–			1.39	1.14 1.69	<0.01	–		
Time since first onset (for every 10 years)	Not selected			–			–			0.53	0.33 0.85	<0.01
HRSD at baseline	Not selected			–			–			–		
Bodily symptoms	0.85	0.73 0.99	0.035	0.85	0.73 0.99	0.036	0.84	0.72 0.98	0.030	0.86	0.74 1.00	0.054
Study level difference												
Duloxetine P3 trial	3.30	1.83 5.96	<0.01	3.40	1.89 6.12	<0.01	3.46	1.92 6.23	<0.01	2.93	1.64 5.23	<0.01
Escitalopram P2 trial	0.84	0.48 1.45	0.525	0.81	0.47 1.40	0.457	0.84	0.48 1.45	0.525	0.79	0.46 1.36	0.398
Escitalopram P3 trial	2.70	1.37 5.33	<0.01	2.47	1.28 4.79	<0.01	2.61	1.33 5.14	<0.01	2.67	1.36 5.25	<0.01
Mirtazapine P3 trial	Ref			Ref			Ref			Ref		

*Two participants in mirtazapine P3 trial were excluded from the final model because they had missing information on 'age at onset'.
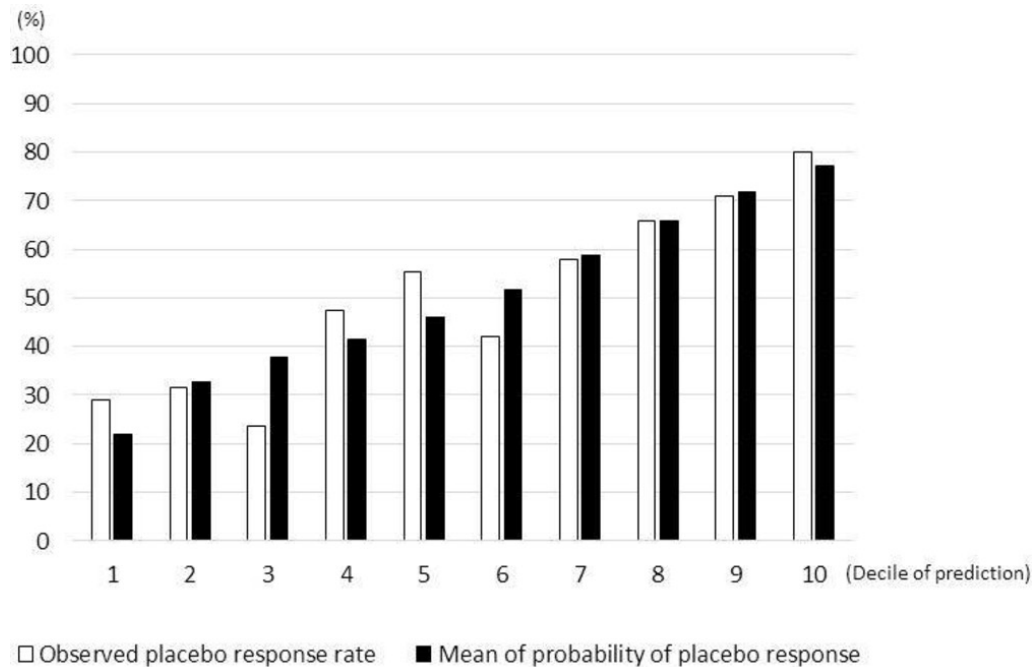HRSD, Hamilton Rating Scale for Depression.

Figure 2 Calibration of the final model. The comparison between observed and predicted placebo response rate in the cohort of four trials.

Performance of the model

Figure 2 and online supplementary table S3 show the calibration of the final model (Hosmer-Lemeshow: p=0.92). Table 4 shows the performance of the final model in terms of discrimination. AUC was 0.70 (95% CI 0.65 to 0.75) when we included all four trials. We also applied the leave-one-trial-out method for the assessment of validity; that is, the model was developed based on three trials, and its validity was examined in the remaining one trial (table 4). The internal validity was variable among different models. AUC was equal to or higher than 0.6 in three models but was only 0.54 when escitalopram P3 trial was used for validation, implying that the discriminatory power of the final model may depend on some study-level factors.

Sensitivity analyses

In the sensitivity analysis using the MI, the same variables were selected with very similar coefficients (online supplementary table S4), and AUC was 0.70 (table 4). The impact of missing values was the same in either completer analysis or MI.

DISCUSSION

We developed a multivariable assessment model for placebo response based on IPD from four randomised trials to identify people who would respond to placebo. Adjusting for differences among trials, three variables emerged as predictors: age at onset, age at baseline and bodily symptoms subscale of HRSD. Time since first onset (time between 'age at onset' and 'age at baseline') had more impact than age itself. Our model suggested individuals with a long history of and severe symptoms were least likely to respond to placebo. There were differences of response rate among studies.

We aimed to develop a model to estimate the probability of placebo response in order to distinguish placebo responders from drug only responders retrospectively in clinical trials. This model was not intended to predict placebo response prospectively or reduce the placebo response rate in clinical trials, but our model is expected to help researchers identify drug-specific effect. By calculating the probability of placebo response using this model, researchers could identify characteristics people who

Table 4 The summary of area under the ROC curve and internal validation				
Method	**Derivation set**	**Validation set**	**Area under curve**	**95% CI**
Internal validation	four trials (completers)	four trials	0.698	0.645 to 0.750
	four trials (multiple imputation)	four trials	0.699	0.649 to 0.750
Leave-one-trial-out cross-validation	Mirtazapine P3 trial, Escitalopram P2 trial, Escitalopram P3 trial	Duloxetine P3 trial	0.655	0.561 to 0.749
	Duloxetine P3 trial, Escitalopram P2 trial, Escitalopram P3 trial	Mirtazapine P3 trial	0.637	0.482 to 0.792
	Duloxetine P3 trial, Mirtazapine P3 trial, Escitalopram P3 trial	Escitalopram P2 trial	0.600	0.469 to 0.733
	Duloxetine P3 trial, Mirtazapine P3 trial, Escitalopram P2 trial	Escitalopram P3 trial	0.541	0.424 to 0.658

ROC, receiver operating characteristic.

would respond to placebo from drug responder group, and put more focus on individuals who would respond only to drug in clinical trials involving neurobiological measurements. Our study also suggested that there is an association between age at onset and placebo response. Age at onset was not often investigated in previous studies and our finding alert that more detail history of illness is needed to develop a prediction model for placebo response. Finally, our model performed slightly better than previous models predicting placebo response.[14 15]

Comparing our model to the two previous models developed by Nelson et al[15] and Nakonezny et al,[14] we found that similar factors were related to placebo response, namely, severity of depression (bodily symptoms), age and age at onset. Nelson et al developed a prediction model of remission to placebo based on eight RCTs with duloxetine for major depression. In their study, they split the dataset of placebo-treated patients and developed the model from the training sample (n=813) and then tested the validity in the test sample (n=204). They identified four factors were associated with remission to placebo: younger age, less severe depression, less anxiety (they developed their subscales and 'anxiety' included items 10, 11, 12, 13 and15 and 17 of HRSD) and shorter current episode. Nakonezny et al also showed baseline severity negatively influences placebo response in adolescents (n=151, 12–17 years of age). Our final model included bodily symptoms (items 11, 13 and 15) of HRSD, and it composed the same anxiety-related physical symptoms as Nelson's anxiety symptoms. Although total HRSD was not selected in our final model, when we inserted total HRSD instead of bodily symptoms, higher total HRSD also related to poor placebo response. Thus, our results do not mean severity of depression is irrelevant to placebo response. We suspect that HRSD was excluded from the final model because bodily symptoms were associated with total HRSD. Overall, our results supported the theory of the relationship between baseline severity, especially anxiety and somatic symptoms, and the placebo response.

However, our results also suggested a reversed association. In their study, younger individuals responded better to placebo, but our study showed older age increases placebo response. However, they did not consider the influence of the age at onset. Our results indicated that three factors, age at onset, baseline age and time since first onset, influence placebo response. Furthermore, time since first onset was a more influential factor to placebo response than other factors. It means that individual with longer time since first onset responds less to placebo than individuals with shorter time since first onset. The association among age-related factors should be investigated in future studies. The performance of our model (AUC=0.70) was slightly better than those of the previous models.[14 15] In their studies, AUC ranged from 0.59 to 0.63 for internal validation. Also, our calibration showed the acceptable performance of the model in Hosmer-Lemeshow test.

There are several limitations to this study. First, because of the limited accessibility of IPD, we included only four trials conducted in Japan. Participants were a highly selected population as they had consented to participate in placebo-controlled clinical trials. The majority of the participants were male and had less severe depression compared with the average of clinical trials.[18] This small sample size and highly selected population can limit the generalisability of our model. Second, we developed a model based on available data. Because of this methodological limitation, we may have missed some important factors in the model. For example, some of the factors that may influence placebo response (eg, precipitating life events and expectation for the treatment) were not measured in our sample and not examined. Third, there was a large variability in placebo response rates (30%–60%)%) among included studies as would generally be expected for antidepressant trials,[1 18] and this study-level difference (ie, baseline risk difference) influences the individual placebo response rate. Baseline risk may differ across the population in prediction model,[27] for various reasons. Although there was no major difference in baseline characteristics of participants in trials (all trials used similar inclusion criteria and were conducted in Japan), there was some heterogeneity among the four trials (placebo run-in, flexible/fixed dosage and sample size), and the baseline risk difference can be derived from the difference in design of the four trials. As we could include only four trials and limited individual information, we were unable to evaluate the influence of design-related predictive factors or other unknown factors. For a comprehensive assessment of both participants and design-related factors and interpret the clinical meaning of predictors, future studies need to include IPD from more trials. However, we believe that this study-level difference would not impair the importance of this model because the individual probability of response to placebo is comparable in the same trial. As we aim to use this model to participants who have already responded to drug, we still can estimate and compare the probability of response to placebo among apparent drug responders in the same trial. Then, we can use the result to distinguish 'drug only responders' from 'placebo responders'. Fourth, our model performance was limited (AUC=0.70 for internal validation), and AUC ranged between 0.54 and 0.66 in the leave-one-trial-out cross-validation. AUCs in leave-one-trial-out cross-validation is generally highly variable (Ensor et al, 2016; Snell et al, 2016) due to various unknown factors. We tried to model heterogeneity among the included studies by including a separate intercept for each study. This was an attempt to account for differences in baseline risks and in other unmeasured variables. Thus far, the performance of our models in the derivation and in the internal cross-validation was similar to many prediction models for various physical illnesses such as seizure recurrence,[28] recurrence of venous thromboembolism[29] and risk of acute kidney injury (Bedford et al, 2016), whose AUC ranged between 0.65[28] and 0.75,[30] and they often showed poorer external validation than internal validation.[31] Therefore, further study is needed to assess the external validity of this model. Lastly, in this study, we used response as our primary outcome. We assumed that placebo response is a stable trait and classified response pattern into response and non-response as many previous models did,[32] because we aimed to identify placebo responders in neurobiological studies. However, we are aware that there is some debate about the reliability and consistency of placebo response.[33] Our binary categorisation can be oversimplistic.

CONCLUSION

Our prediction model to estimate the probability of responding to placebo in antidepressant trials can help researchers distinguish individuals who would have responded to placebo among apparent drug responders. Its moderate performance is slightly better than previous placebo-response prediction models, but internal validity was variable. Our model suggested that age, age at onset and severity of depression may be related to placebo response, partly replicating previous models. However, more detailed individual participant information from large number of studies is needed to examine all relevant predictive factors. Also, the external validity of this model should be examined in an independent sample.

REFERENCES

1. Furukawa TA, Cipriani A, Atkinson LZ, *et al*. Placebo response rates in antidepressant trials: a systematic review of published and unpublished double-blind randomised controlled studies. *Lancet Psychiatry* 2016;3:1059–66.
2. Furukawa TA, Cipriani A, Leucht S, *et al*. Is placebo response in antidepressant trials rising or not? A reanalysis of datasets to conclude this long-lasting controversy. *Evid Based Ment Health* 2018;21:1–3.
3. Khan A, Detke M, Khan SR, *et al*. Placebo response and antidepressant clinical trial outcome. *J Nerv Ment Dis* 2003;191:211–8.
4. Weimer K, Colloca L, Enck P. Placebo eff ects in psychiatry: mediators and moderators. *Lancet Psychiatry* 2015;2:246–57.
5. Borsook D, Becerra L, Fava M. Use of functional imaging across clinical phases in CNS drug development. *Transl Psychiatry* 2013;3:e282.
6. Muthén B, Brown HC. Estimating drug effects in the presence of placebo response: causal inference using growth mixture modeling. *Stat Med* 2009;28:3363–85.
7. 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 Psychiatry* 2015;72:1087–94.
8. Sikora M, Heffernan J, Avery ET, *et al*. Salience network functional connectivity predicts placebo effects in major depression. *Biol Psychiatry Cogn Neurosci Neuroimaging* 2016;1:68–76.
9. Khan A, Brown WA. Antidepressants versus placebo in major depression: an overview. *World Psychiatry* 2015;14:294–300.
10. Rutherford BR, Roose SP. A model of placebo response in antidepressant clinical trials. *Am J Psychiatry* 2013;170:723–33.
11. Papakostas GI, Østergaard SD, Iovieno N. The nature of placebo response in clinical studies of major depressive disorder. *J Clin Psychiatry* 2015;76:456–66.
12. Salanti G, Chaimani A, Furukawa TA, *et al*. Impact of placebo arms on outcomes in antidepressant trials: systematic review and meta-regression analysis. *Int J Epidemiol* 2018;47:1454–64.
13. Furukawa TA, Weitz ES, Tanaka S, *et al*. Initial severity of depression and efficacy of cognitive-behavioural therapy: individual-participant data meta-analysis of pill-placebo-controlled trials. *Br J Psychiatry* 2017;210:190–6.
14. Nakonezny PA, Mayes TL, Byerly MJ, *et al*. Predicting placebo response in adolescents with major depressive disorder: The Adolescent Placebo Impact Composite Score (APICS). *J Psychiatr Res* 2015;68:346–53.
15. Nelson JC, Zhang Q, Deberdt W, *et al*. Predictors of remission with placebo using an integrated study database from patients with major depressive disorder. *Curr Med Res Opin* 2012;28:325–34.
16. Moons KG, Altman DG, Reitsma JB, *et al*. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015;162:W1–73.
17. Furukawa TA, Maruo K, Noma H, *et al*. Initial severity of major depression and efficacy of new generation antidepressants: individual participant data meta-analysis. *Acta Psychiatr Scand* 2018;137:450–8.
18. Cipriani A, Furukawa TA, Salanti G, *et al*. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *Lancet* 2018;391:1357–66.
19. Higuchi T, Murasaki M, Kamijima K. Clinical evaluation of duloxetine in the treatment of major depressive disorder. *Jpn J Clin Psychopharmacol* 2009;12:1613–34.
20. A double-blind KT. placebocontrolled study of a new antidepressant, mirtazapine, in depressed patients. *Jpn J Clin Psychopharmacol* 2009;12:289–306.
21. Hirayasu Y. A dose-response and non-inferiority study evaluating the efficacy and safety of escitalopram inpatients with major depressive disorder. *Jpn J Clin Psychopharmacol* 2011;14:883–99.
22. Hirayasu Y. A dose-response study of escitalopram in patients with major depressive disorder: a placebo-controlled, double-blind study. *Jpn J Clin Psychopharmacol* 2011;14:871–82.
23. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry* 1960;23:56–62.
24. Kessler RC, van Loo HM, Wardenaar KJ, *et al*. Using patient self-reports to study heterogeneity of treatment effects in major depressive disorder. *Epidemiol Psychiatr Sci* 2017;26:22–36.
25. Furukawa TA, Streiner DL, Azuma H, *et al*. Cross-cultural equivalence in depression assessment: Japan-Europe-North American study. *Acta Psychiatr Scand* 2005;112:279–85.
26. Sterne JA, White IR, Carlin JB, *et al*. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* 2009;338:b2393.
27. Snell KI, Hua H, Debray TP, *et al*. Multivariate meta-analysis of individual participant data helped externally validate the performance and implementation of a prediction model. *J Clin Epidemiol* 2016;69:40–50.
28. Lamberink HJ, Boshuisen K, Otte WM, *et al*. Individualized prediction of seizure relapse and outcomes following antiepileptic drug withdrawal after pediatric epilepsy surgery. *Epilepsia* 2018;59:e28–33.
29. Ensor J, Riley RD, Jowett S, *et al*. Prediction of risk of recurrence of venous thromboembolism following treatment for a first unprovoked venous thromboembolism: systematic review, prognostic model and clinical decision rule, and economic evaluation. *Health Technol Assess* 2016;20:1–190.
30. Bedford M, Stevens P, Coulton S, *et al*. Development of risk models for the prediction of new or worsening acute kidney injury on or during hospital admission: a cohort and nested study. *Health Services and Delivery Research* 2016;4:1–160.
31. Ohnuma T, Uchino S. Prediction Models and Their External Validation Studies for Mortality of Patients with Acute Kidney Injury: A Systematic Review. *PLoS One* 2017;12:e0169341.
32. Kessels R, Mozer R, Bloemers J. Methods for assessing and controlling placebo effects. *Stat Methods Med Res* 2017;962280217748339:096228021774833.
33. Kaptchuk TJ, Kelley JM, Deykin A, *et al*. Do "placebo responders" exist? *Contemp Clin Trials* 2008;29:587–95.