ORIGINAL RESEARCH

Sex differences in the prediction of the effectiveness of paroxetine for patients with major depressive disorder identified using a receiver operating characteristic curve analysis for early response

Tetsu Tomita
Yasui-Furukori Norio
Yasushi Sato
Taku Nakagami
Shoko Tsuchimine
Ayako Kaneda
Sunao Kaneko

Department of Neuropsychiatry,
Graduate School of Medicine,
Hirosaki University, Hirosaki, Japan

Correspondence: Tetsu Tomita
Department of Neuropsychiatry,
Graduate School of Medicine,
Hirosaki University, 5 Zaifu-cho,
Hirosaki, 036-8562, Japan
Tel +81 172 39 5066
Fax +81 172 39 5067
Email ttomita1984@yahoo.co.jp

Background: We investigated cutoff values for the early response of patients with major depressive disorder to paroxetine and their sex differences by using a receiver operating characteristic (ROC) curve analysis to predict the effectiveness of paroxetine.

Methods: In total, 120 patients with major depressive disorder were enrolled and treated with 10–40 mg/day paroxetine for 6 weeks; 89 patients completed the protocol. A clinical evaluation using the Montgomery–Asberg Depression Rating Scale (MADRS) was performed at weeks 0, 1, 2, 4, and 6.

Results: In male subjects, the cutoff values for MADRS improvement rating in week 1, week 2, and week 4 were 20.9%, 34.9%, and 33.3%, respectively. The sensitivities and the specificities were 83.3% and 80.0%, 83.3% and 80.0%, and 100% and 90%, respectively. The areas under the curve (AUC) were 0.908, 0.821, and 0.979, respectively. In female subjects, the cutoff values for the MADRS improvement rating in week 1, week 2, and week 4 were 21.4%, 35.7%, and 32.3%, respectively. The sensitivities and the specificities were 71.4% and 84.6%, 73.8% and 76.9%, and 90.5% and 76.9%, respectively. The AUCs were 0.781, 0.735, and 0.904, respectively.

Conclusion: Early improvement with paroxetine may predict the long-term response. The accuracy of the prediction for the response is higher in male subjects.

Keywords: antidepressants, paroxetine, early response, sex differences, receiver operating characteristic curve analysis

Introduction

While many researchers have tried to optimize the pharmacological treatment of patients with major depressive disorder (MDD), the efficacy and tolerability of the medications prescribed remains highly variable. In addition to clinical heterogeneity and diagnostic uncertainty,[1–3] environmental,[4] social, and genetic factors can contribute to individual variation in the therapeutic or toxic effects of antidepressants.[5,6] Previous studies have reported that neurofunctional measurement using electroencephalography has the potential to predict patient responses to antidepressant treatment.[7–9]

Some previous research has shown sex-specific differences in the response to antidepressants in patients with MDD. Khan et al reported that female patients with MDD had a significantly higher response than men to selective serotonin reuptake inhibitor (SSRI) antidepressants.[10] Morishita and Arita, and Morishita and Kinoshita reported

that female patients with MDD had a significantly higher response than males to sertraline, and male patients with MDD had a higher response than females to milnacipran.[11,12] These studies demonstrate that sex differences can affect the efficacy of antidepressant treatment in patients with MDD.

Recently, some studies have suggested that the response to an antidepressant in the early phase of treatment for patients with MDD predicts the long-term effect of the antidepressant and is important in determining whether the prescription should be changed. Inagaki et al reported that the rate of reduction in the Hamilton Rating Scale for Depression (HAM-D) on the third day from the administration of paroxetine for patients with MDD was higher in responders to the paroxetine treatment than it was in nonresponders.[13] Nakajima et al demonstrated the efficacy of switching antidepressants early in the treatment of patients who failed to respond to the initial antidepressants used. When patients with MDD failed to respond to sertraline at the 2nd week from baseline, the patients whose sertraline was switched to paroxetine showed higher responder and remitter rates for the endpoint than patients whose sertraline was not changed.[14]

Receiver operating characteristic (ROC) curve analysis is used to analyze various cutoff values and is helpful for predicting the treatment response for patients with psychiatric disorders. Some studies using ROC curve analysis have suggested the level of improvement in symptoms of MDD that should be considered as the antidepressant response predictor in the early treatment stages.[15–17] Kok et al suggested that the treatment should be changed if, after 3–4 weeks, less than a 30% improvement in the depression scale score (HAM-D and the Montgomery-Asberg Depression Rating Scale [MADRS]) is achieved in elderly patients with MDD.[16] Henkel et al showed that a 20% reduction in the HAM-D baseline score at day 14 relative to the baseline could predict the response to treatment with antidepressants and remission from MDD.[15] With regards to individual antidepressants, Lin et al studied patients with MDD who began their treatment with fluoxetine. At weeks 1, 2, 3, and 4 from the baseline, a HAM-D 17 score reduction of 25%, 39%, 43%, and 50% seemed to be the potential cutoff to predict the response to fluoxetine.[17] However, no studies have addressed the degree of improvement in MDD symptoms that should be considered as the treatment response predictor in the early stages of treatment with paroxetine.

In the present study, we analyzed the cutoff values for the early response to paroxetine in MDD by using a ROC curve analysis to predict the effectiveness of paroxetine. In addition, we analyzed the ROC curve separately for male and female subjects and compared the results.

Methods
Patients
Patients eligible for this study included male and female patients aged 18–70 years. They suffered from MDD according to the Diagnostic and Statistical Manual of Mental Disorders Fourth Edition and they were required to score more than 20 points on the MADRS.[18] The MADRS consists of ten items and is scored on a scale ranging from 0–6 for each item. The patients were required to be medication-free, including the absence of any psychotropic agents, for at least 1 month prior to the start of the study. Patients were excluded if they had clinically significant abnormal electrocardiography or laboratory findings; a past history of mental disorder other than MDD, for example, bipolar affective disorder, schizophrenia, alcoholism, epilepsy, or drug abuse; or any clinically significant organic or neurological disease.

In total, 120 patients were enrolled in the study upon admission, and 89 patients completed the study (males:females =34:55). The mean ± standard deviation age was 46.7±13.5 years, and the mean body weight was 55.8.0±10.5 kg. In total, 31 patients withdrew from the present study for the following reasons: 17 experienced severe side effects, four took drugs not permitted by the study protocol, seven did not come to the hospital for unknown reasons, and three withdrew their consent for personal reasons.

This study was approved by the Ethics Committee of Hirosaki University Hospital, and written informed consent to participate in this study was obtained from either the patients or their authorized representatives before the study.

Protocol
The assessment of pretreatment clinical status using the structured interview from the MADRS (a structured interview guide for the MADRS [SIGMA]) for depressive symptoms and the Udvalg for Klinicke Undersogelser (UKU) for side effects was performed by two well-trained psychiatrists.[19,20] The UKU is a comprehensive scale for the side effects of psychotropic drugs and consists of 48 items rated from zero to three according to the presence or severity of the side effects.

For the first week, a dose of 20 mg/day paroxetine (Paxil®; Glaxo-SmithKline plc, Brentford, England) was administered at 8.00 pm; thereafter, the dose of paroxetine was increased to 40 mg/day from week 2 to week 6. Paroxetine has a simple function and mainly modulates serotoninergic function.[21]

We chose paroxetine to treat the participants because we thought that there might be less influence on other functions, and paroxetine has been widely used to treat patients with MDD. The dose was maintained if mild side effects (1 point in the UKU) were observed. The dose of paroxetine was decreased if moderate side effects were observed (two points in the UKU), and administration was discontinued in cases of severe side effects (three points in the UKU). No other drugs were given except diazepam (2–5 mg/day, n=19) for anxiety symptom, brotizolam (0.25 mg/day, n=20; 4 mg/day, n=17) for insomnia, and sennoside (12–48 mg/day, n=12) as a laxative for constipation. Clinical symptoms were evaluated at weeks 1, 2, 4, and 6 of treatment.

Data analyses and statistics

We defined responders as patients with MADRS improvements of >50% over the baseline score at week 6 of treatment. This procedure has been widely used in previous studies of the response to paroxetine or antidepressants.[14,22,23] A Student's t-test and a chi-squared test were performed to compare demographic data and MADRS scores between responders and non-responders and between male and female participants. We used an ROC curve to analyze the MADRS improvement rating to determine the cutoff points that yield the highest combined sensitivity and specificity for distinguishing responders and nonresponders among all of the participants and also among males and females. $P<0.05$ was regarded as statistically significant. All analyses were performed using SPSS version 20 for Windows (IBM Corporation, Armonk, NY, USA) and StatFlex version 6 (Artech Co, Ltd, Tokyo, Japan).

Results
Demographic and clinical characteristics

The demographic and clinical data of the subjects (total subjects, responders, and nonresponders) are listed in Table 1. There were significant differences in the MADRS scores at week 6 and the MADRS improvement rating between the responders and nonresponders according to Student's t-test. There was no significant difference in sex between responders and nonresponders according to the chi-squared test.

Table 2 shows the demographic and clinical data of the male and female subjects. There was a significant difference from the baseline only in the MADRS score in week 1. The female subjects showed a significantly higher MADRS score than did the male subjects in week 1. There were no significant differences in other clinical indices.

Table 1 Demographic and clinical data on the subjects and the results of Student's t-test and the chi-squared test between responders and nonresponders

	Total (n=89)	Responders (n=66)	Nonresponders (n=23)	P-value
Age	46.7±13.5	48.0±12.8	43.0±15.1	0.169
Sex (M:F)	34:55	24:42	10:13	0.545
MADRS score				
W0	39.5±9.2	39.1±9.0	40.4±10.0	0.599
W1	28.6±11.1	26.1±10.8	35.9±8.5	0.000*
W2	21.3±12.4	18.6±11.2	29.0±12.9	0.002*
W4	16.6±12.5	12.0±9.6	29.9±10.4	0.000*
W6	14.1±13.1	8.3±7.9	30.6±11.0	0.000*
MADRS improvement rate (%)				
W1	26.3±30.0	32.5±30.8	8.7±18.8	0.000*
W2	46.7±27.7	52.6±26.1	30.0±25.8	0.001*
W4	58.6±28.3	70.2±20.9	25.4±18.6	0.000*
W6	64.9±30.3	79.0±18.4	24.2±18.2	0.000*

Note: *Statistically significant value ($P<0.01$).
Abbreviations: MADRS, Montgomery-Asberg Depression Rating Scale; W, week.

ROC curve analysis for all subjects

Figure 1 shows the ROC curves for the response in all subjects. The threshold for the response that gave the maximal sensitivity and specificity for the MADRS improvement rating from baseline in the week 1 was 21.4%. The threshold for the response that gave the maximal sensitivity and specificity for the MADRS improvement rating from baseline in week 2 was 35.7%. The threshold for response that gave the maximal sensitivity and specificity for the MADRS improvement rating from baseline in week 4 was 33.3%. There was a significant difference in the area under the curve (AUC) between weeks 2 and 4 ($P=0.014$).

Table 2 Demographic and clinical data on the male and female subjects

	Male (n=34)	Female (n=55)	P-value
Age	46.2±11.5	47.0±14.8	0.791
Responder (%)	24 (70.6%)	42 (76.4%)	0.545
MADRS score			
W0	37.5±8.4	40.7±9.6	0.109
W1	24.7±10.5	31.0±10.8	0.008*
W2	18.2±10.9	23.3±13.0	0.062
W4	16.0±10.6	17.0±13.7	0.711
W6	14.8±10.3	13.6±14.7	0.684
MADRS improvement rate (%)			
W1	33.0±27.7	22.2±30.8	0.099
W2	52.5±25.8	43.2±28.5	0.124
W4	56.4±27.4	60.0±29.0	0.558
W6	59.5±27.0	68.2±31.9	0.187

Note: *Statistically significant value ($P<0.01$).
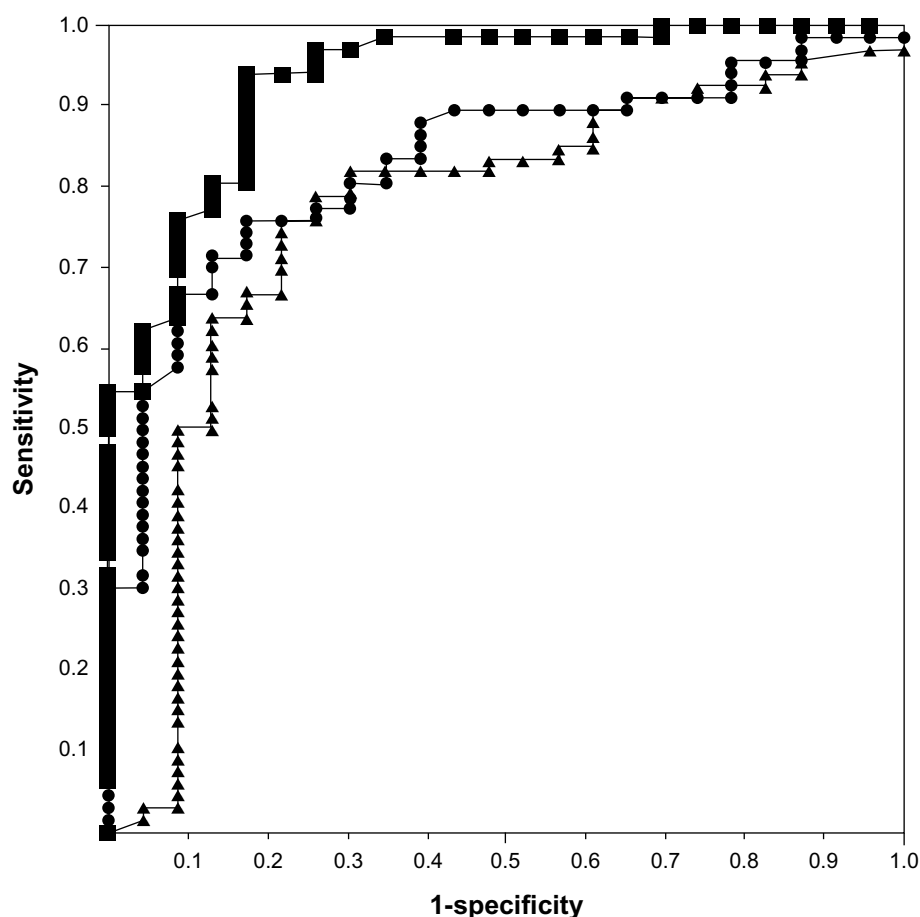**Abbreviations:** MADRS, Montgomery-Asberg Depression Rating Scale; W, week.

Figure 1 ROC curves for all subjects.
Notes: ROC curves for the response at weeks 1, 2, and 4 for all subjects to determine the cutoff points for the MADRS improvement rating. The circles show the ROC curve for week 1 (AUC =0.829), the triangles show the ROC curve for week 2 (AUC =0.764), and the squares show the ROC curve for week 4 (AUC =0.932).
Abbreviations: ROC, receiver operating characteristic; MADRS, Montgomery–Asberg Depression Rating Scale; AUC, area under curve.

ROC curve analysis for male subjects and female subjects

Figure 2 shows the ROC curves in male subjects in addition to the fraction of true positive results and false negative results for various cutoff levels for the MADRS improvement rating from baseline at weeks 1, 2, and 4 for the response to paroxetine in week 6. The threshold for response that gave the maximal sensitivity and specificity for the MADRS improvement rating from the baseline in week 2 was 20.9%. The threshold for response that gave the maximal sensitivity and specificity for the MADRS improvement rating from baseline in week 2 was 34.9%. The threshold for response that gave the maximal sensitivity and specificity for the MADRS improvement rating from baseline in week 4 was 33.3%. There were no differences in the AUC among weeks 1, 2, and 4.

Figure 3 shows the ROC curves in female subjects in addition to the fraction of true positive results and false negative

results for various cutoff levels of the MADRS improvement rating from baseline in the 1st, 2nd, and 4th weeks for the response to paroxetine in the 6th week. The threshold for response that gave the maximal sensitivity and specificity for the MADRS improvement rating from baseline in the 1st week was 21.4%. The threshold for response that gave the maximal sensitivity and specificity for the MADRS improvement rating from baseline in the 2nd week was 35.7%. The threshold for response that gave the maximal sensitivity and specificity for the MADRS improvement rating from baseline in week 4 was 32.3%. There were no differences in the AUC among weeks 1, 2, and 4.

Table 3 summarizes the results of the ROC curve analysis for all of the subjects.

Discussion

In the present study, the MADRS improvement rating in the early treatment phase was found to predict the long-term
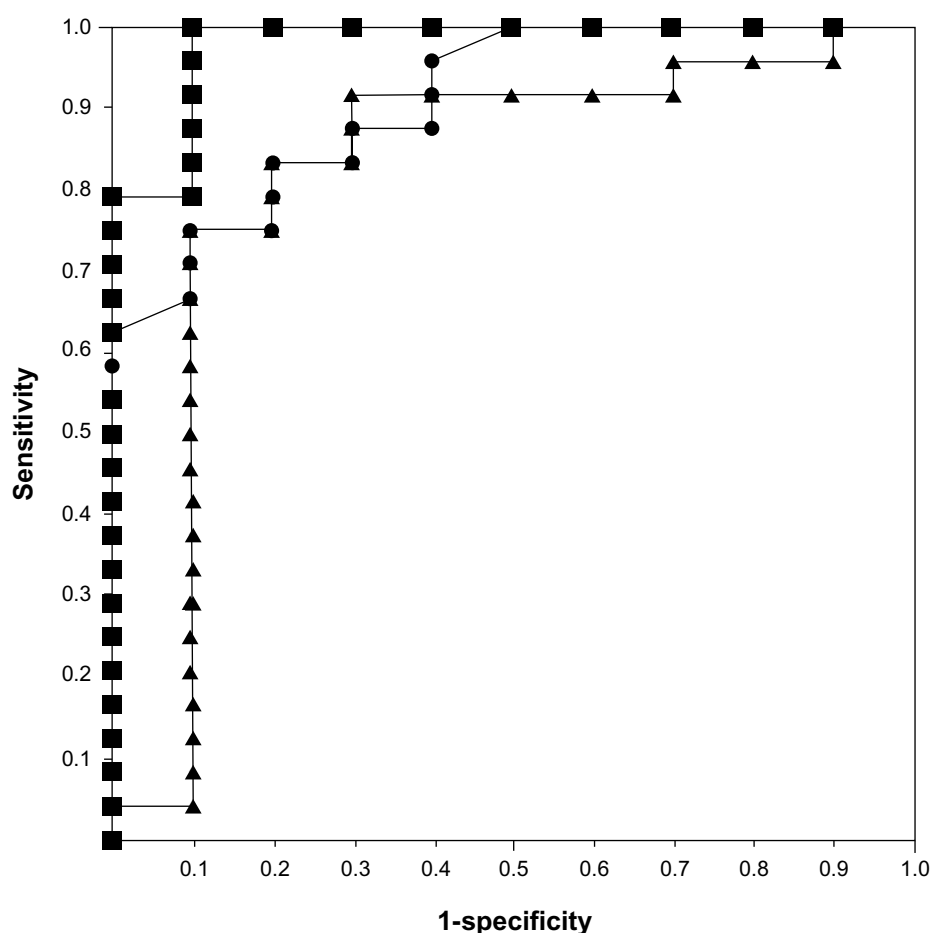
Figure 2 ROC curves in male subjects.
Notes: ROC curves for response at weeks 1, 2, and 4 in the male subjects to determine the cutoff points for the MADRS improvement rating. The circles show the ROC curve for week 1 (AUC =0.908), the triangles show the ROC curve for week 2 (AUC =0.821), and the squares show the ROC curve for week 4 (AUC =0.979).
Abbreviations: ROC, receiver operating characteristic; MADRS, Montgomery-Asberg Depression Rating Scale; AUC, area under curve.

response to the antidepressant; the accuracy of the prediction was also found to be higher in male subjects. This study is the first to report sex differences in the prediction of the effectiveness of antidepressants for patients with MDD by using an ROC curve analysis for early response. This is also the first study to investigate the cutoff values for MADRS improvement in the early stages of MDD treatment with paroxetine to predict the endpoint response to the treatment.

Our results are highly relevant to the clinical setting. Until now, it has been considered that antidepressants should be used at a sufficient dose and for a sufficient duration for patients with MDD, and that a long duration is required to determine whether the antidepressant is effective.[24] According to our results, only a few weeks may be sufficient to evaluate the efficacy of the antidepressant and to decide whether to change the prescription. The introduction of our findings into clinical decisions may reduce the period during which

patients suffer from MDD symptoms. Our results support those of Nakajima et al[14]; they recommended switching the antidepressant in the early phase of the initial treatment if the patient does not respond well. The present study is preliminary and needs to be followed up, but a cutoff value of approximately 20%–35% improvement in MDD symptoms from baseline should typically be considered when deciding whether patients with MDD respond to the initial antidepressant in the early phase and whether the prescription should be changed.

In a previous study investigating cutoff values for depression scale improvement in the early treatment stage of patients with MDD by ROC curve analysis, cutoff values ranging from 20%–39% in weeks 1 or 2 from baseline were found to be useful in predicting the response to the antidepressant. The previous studies used antidepressants other than paroxetine; however, our results are similar to those of previous studies. It may be possible to apply the cutoff value of 20%–39%
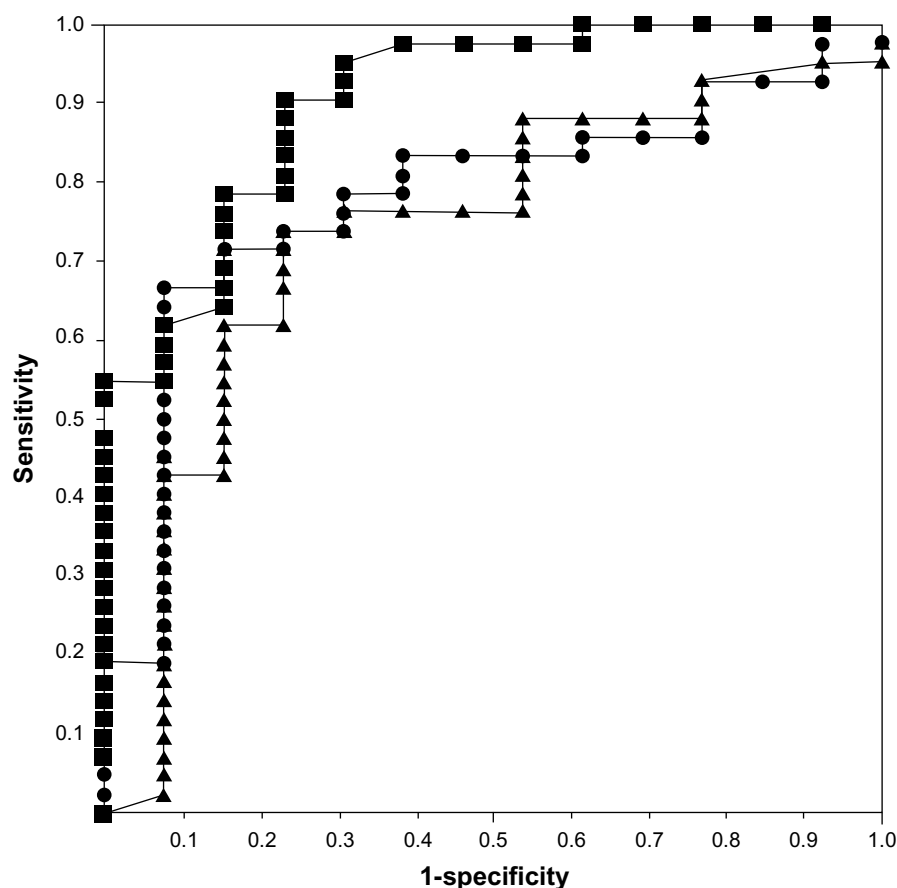
Figure 3 ROC curves in female subjects.
Notes: ROC curves for response at weeks 1, 2, and 4 in the female subjects to determine the cutoff points for the MADRS improvement rate. The circles show the ROC curve for week 1 (AUC =0.781), the triangles show the ROC curve for week 2 (AUC =0.735), and the squares show the ROC curve for week 4 (AUC =0.904).
Abbreviations: ROC, receiver operating characteristic; MADRS, Montgomery-Asberg Depression Rating Scale; AUC, area under curve.

improvement in the early phase to the treatment of patients with MDD using other antidepressants.

It is unclear why the MADRS improvement rating predicts the response better for male participants than for female participants, but previous studies have reported sex

differences in the response to SSRIs among patients with MDD. Kornstein et al reported that female patients with MDD showed the best responses to SSRIs, whereas male patients with MDD showed the best responses to tricyclic antidepressants (TCA).[25] In studies investigating the

Table 3 The results of ROC curve analysis for total subjects (male and female subjects)

Predictors	Cut off (%)	AUC	Sensitivity (%)	Specificity (%)	Odds ratio	Likelihood ratio
MADRS improvement rate of total subjects						
week 1	21.4	0.829	75.8	82.6	14.8	4.36
week 2	35.7	0.764	75.8	78.3	11.3	3.48
week 4	33.3	0.932	93.9	82.6	73.6	5.40
MADRS improvement rate of male subjects						
week 1	20.9	0.908	83.3	80.0	20.0	4.17
week 2	34.9	0.821	83.3	80.0	20.0	4.17
week 4	33.3	0.979	100	90.0	∞	10.0
MADRS improvement rate of female subjects						
week 1	21.4	0.781	71.4	84.6	13.8	4.64
week 2	35.7	0.735	73.8	76.9	9.4	3.20
week 4	32.3	0.904	90.5	76.9	31.7	3.92

Abbreviations: ROC, receiver operating characteristic; MADRS, Montgomery-Asberg Depression Rating Scale; AUC, area under curve.

response to SSRI or citalopram, female patients with MDD had a better response to these drugs than male patients with MDD.[10,26] Some studies have reported epidemiological sex differences. Females are more prone to MDD and recurrent depression and are more likely to develop MDD as a response to stress.[27–29] Personality traits differ between males and females,[30] and such personality characteristics may affect the response to the antidepressant treatment. In a study using positron emission tomography for healthy subjects, female subjects had significantly higher 5-HT$_{1A}$ receptor binding potentials and lower 5-HTT binding potentials than male subjects in many brain regions[31]; female subjects also had higher 5-HT$_{1A}$ binding potentials than male subjects.[32] Bethea et al suggested that the presence of estrogen increases in serotonin transporter expression in the hypothalamus.[33] Such biological differences and hormonal fluctuations in females may have been partly responsible for the sex differences in the present study.

There are some limitations in the present study. First, the results of the present study do not indicate whether the initial antidepressant should be changed if the MADRS improvement rating does not reach 20%–35%. The present study is not an intervention study and we did not confirm the clinical significance of changing the initial antidepressants when the MADRS improvement was 20%–35% over the baseline. Nakajima et al studied the switching of sertraline according to early responses; our study would have benefitted from such an intervention design.[14] Second, it is unclear when the initial antidepressants should be switched according to the MADRS improvement rating. We showed that the response in week 1 and week 2 could predict the response to treatment, but the response in week 4 could also predict the response. Therefore, it is possible that the initial antidepressant should not be switched until it has been used for over 4 weeks. Third, the patients' ages spanned a large range. Several previous studies showed that age was related to the clinical response and influenced which antidepressants were most effective for the individual;[34,35] thus, age may be an important factor in the treatment of MDD. In addition, the sample size was small and the dropout rate was comparatively large in the study. These variables may have affected the validity of our results. Fourth, there might be several statistical problems. The sex differences in the AUC might be brought about by the lack of a significant difference in the baseline MADRS score. If the baseline MADRS scores were more uniform, the results might change. Furthermore, circularity or collinearity might exist in the ROC analysis because we used the MADRS for the early response and the final response.

Conclusion

In conclusion, we showed that the MADRS improvement rating in the early phase can predict the response to paroxetine and that the accuracy of prediction is higher in male subjects according to the ROC curve analysis for the early response. However, further clinical studies are necessary to apply these results from the present study into the clinical setting.

Acknowledgments

Disclosure

The authors report no conflicts of interest in this work.

References

1. Antonijevic IA. Depressive disorders – is it time to endorse different pathophysiologies? *Psychoneuroendocrinology*. 2006;31(1):1–15.
2. van Praag HM. Kraepelin, biological psychiatry, and beyond. *Eur Arch Psychiatry Clin Neurosci*. 2008;258 Suppl 2:29–32.
3. Lamers F, de Jonge P, Nolen WA, et al. Identifying depressive subtypes in a large cohort study: results from the Netherlands Study of Depression and Anxiety (NESDA). *J Clin Psychiatry*. 2010;71(12): 1582–1589.
4. Abela JR, Skitch SA, Auerbach RP, Adams P. The impact of parental borderline personality disorder on vulnerability to depression in children of affectively ill parents. *J Pers Disord*. 2005;19(1):68–83.
5. Arias B, Gutiérrez B, Pintor L, Gastó C, Fañanás L. Variability in the 5-HT(2A) receptor gene is associated with seasonal pattern in major depression. *Mol Psychiatry*. 2001;6(2):239–242.
6. Caspi A, Sugden K, Moffitt TE, et al. Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. *Science*. 2003;301(5631):386–389.
7. Cook IA, Hunter AM, Abrams M, Siegman B, Leuchter AF. Midline and right frontal brain function as a physiologic biomarker of remission in major depression. *Psychiatry Res*. 2009;174(2):152–157.
8. Jaworska N, Protzner A. Electrocortical features of depression and their clinical utility in assessing antidepressant treatment outcome. *Can J Psychiatry*. 2013;58(9):509–514.
9. Jaworska N, De Somma E, Blondeau C, et al. Auditory P3 in antidepressant pharmacotherapy treatment responders, nonresponders and controls. *Eur Neuropsychopharmacol*. 2013;23(11): 1561–1569.
10. Khan A, Brodhead AE, Schwartz KA, Kolts RL, Brown WA. Sex differences in antidepressant response in recent antidepressant clinical trials. *J Clin Psychopharmacol*. 2005;25(4):318–324.
11. Morishita S, Arita S. Differential effects of milnacipran, fluvoxamine and paroxetine for depression, especially in gender. *Eur Psychiatry*. 2003;18(8):418–420.
12. Morishita S, Kinoshita T. Predictors of response to sertraline in patients with major depression. *Hum Psychopharmacol*. 2008;23(8): 647–651.
13. Inagaki T, Furuya M, Kawamukai T, et al. Prediction of response within the first 3 days to treatment with paroxetine for depression. *Prim Care Companion J Clin Psychiatry*. 2008;10(2):129–132.
14. Nakajima S, Uchida H, Suzuki T, et al. Is switching antidepressants following early nonresponse more beneficial in acute-phase treatment of depression?: a randomized open-label trial. *Prog Neuropsychopharmacol Biol Psychiatry*. 2011;35(8):1983–1989.

15. Henkel V, Seemüller F, Obermeier M, et al. Does early improvement triggered by antidepressants predict response/remission? Analysis of data from a naturalistic study on a large sample of inpatients with major depression. *J Affect Disord*. 2009;115(3):439–449.

16. Kok RM, van Baarsen C, Nolen WA, Heeren TJ. Early response as predictor of final remission in elderly depressed patients. *Int J Geriatr Psychiatry*. 2009;24(11):1299–1303.

17. Lin CH, Lane HY, Chen CC, Juo SH, Yen CF. Early prediction of fluoxetine response for Han Chinese inpatients with major depressive disorder. *J Clin Psychopharmacol*. 2011;31(2):187–193.

18. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry*. 1979;134:382–389.

19. Takahashi N, Tomita K, Higuchi T, Inada T. The inter-rater reliability of the Japanese version of the Montgomery-Asberg depression rating scale (MADRS) using a structured interview guide for MADRS (SIGMA). *Hum Psychopharmacol*. 2004;19(3):187–192.

20. Lingjaerde O, Ahlfors UG, Bech P, Dencker SJ, Elgen K. The UKU side effect rating scale. A new comprehensive rating scale for psychotropic drugs and a cross-sectional study of side effects in neuroleptic-treated patients. *Acta Psychiatr Scand Suppl*. 1987;334:1–100.

21. Burrows GD, McIntyre IM, Judd FK, Norman TR. Clinical effects of serotonin reuptake inhibitors in the treatment of depressive illness. *J Clin Psychiatry*. 1988;49 Suppl:18–22.

22. Trivedi MH, Corey-Lisle PK, Guo Z, Lennox RD, Pikalov A, Kim E. Remission, response without remission, and nonresponse in major depressive disorder: impact on functioning. *Int Clin Psychopharmacol*. 2009;24(3):133–138.

23. Blier P, Gobbi G, Turcotte JE, et al. Mirtazapine and paroxetine in major depression: a comparison of monotherapy versus their combination from treatment initiation. *Eur Neuropsychopharmacol*. 2009;19(7):457–465.

24. Practice guideline for the treatment of patients with major depressive disorder (revision). American Psychiatric Association. *Am J Psychiatry*. 2000;157(Suppl 4):1–45.

25. Kornstein SG, Schatzberg AF, Thase ME, et al. Gender differences in treatment response to sertraline versus imipramine in chronic depression. *Am J Psychiatry*. 2000;157(9):1445–1452.

26. Young EA, Kornstein SG, Marcus SM, et al. Sex differences in response to citalopram: a STAR*D report. *J Psychiatr Res*. 2009;43(5):503–511.

27. Ernst C, Angst J: The Zurich Study. XII: Sex differences in depression. Evidence from longitudinal epidemiological data. *Eur Arch Psychiatry Clin Neurosci*. 1992;241:222–230.

28. Kendler KS, Thornton LM, Gardner CO. Stressful life events and previous episodes in the etiology of major depression in women: an evaluation of the "kindling" hypothesis. *Am J Psychiatry*. 2000;157(8): 1243–1251.

29. Kessler RC, McGonagle KA, Swartz M, Blazer DG, Nelson CB. Sex and depression in the National Comorbidity Survey. I: Lifetime prevalence, chronicity and recurrence. *J Affect Disord*. 1993;29(2–3):85–96.

30. Del Giudice M, Booth T, Irwing P. The distance between Mars and Venus: measuring global sex differences in personality. *PLoS One*. 2012;7(1):e29265.

31. Jovanovic H, Lundberg J, Karlsson P, et al. Sex differences in the serotonin 1A receptor and serotonin transporter binding in the human brain measured by PET. *Neuroimage*. 2008;39(3):1408–1419.

32. Parsey RV, Oquendo MA, Simpson NR, et al. Effects of sex, age, and aggressive traits in man on brain serotonin 5-HT1A receptor binding potential measured by PET using [C-11]WAY-100635. *Brain Res*. 2002;954(2):173–182.

33. Bethea CL, Lu NZ, Gundlah C, Streicher JM. Diverse actions of ovarian steroids in the serotonin neural system. *Front Neuroendocrinol*. 2002;23(1):41–100.

34. Higuchi H, Sato K, Naito S, et al. Differential clinical effects of fluvoxamine by the effect of age in Japanese female major depressive patients. *Neuropsychiatr Dis Treat*. 2009;5:151–155.

35. Morishita S, Arita S. Differential effects of fluvoxamine, paroxetine and milnacipran for depression, especially with regard to age. *Hum Psychopharmacol*. 2004;19(6):405–408.