Published in final edited form as:

J Psychiatr Res. 2016 October ; 81: 112–118. doi:10.1016/j.jpsychires.2016.07.005.

Combining moderators to identify clinical profiles of patients who will, and will not, benefit from aripiprazole augmentation for treatment resistant late-life major depressive disorder

Stephen F. Smagula¹, Meredith L. Wallace¹, Stewart J. Anderson², Jordan F. Karp^{1,3}, Eric J. Lenze⁴, Benoit H. Mulsant⁵, Meryl A. Butters¹, Daniel M. Blumberger⁵, Breno S. Diniz⁶, Francis Lotrich¹, Mary Amanda Dew^{1,7}, and Charles F. Reynolds III^{1,8}

¹Department of Psychiatry, Western Psychiatric Institute and Clinic of University of Pittsburgh Medical Center, Pittsburgh, PA, USA

²Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA

³VAPHS, Geriatric Research, Education, and Clinical Center

⁴Healthy Mind Lab, Department of Psychiatry, Washington University School of Medicine, St Louis, MO, USA

⁵Centre for Addiction and Mental Health, and Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada

Corresponding author: Stephen F. Smagula, PhD, Department of Psychiatry, Western Psychiatric Institute and Clinic, University of Pittsburgh Medical Center, 3811 O'Hara St, Pittsburgh, PA 15217, sfs26@pitt.edu.

Contributors: All authors contributed to drafting and critically revising the manuscript. SFS and MLW conducted the analysis. SJA, JFK, EJJ, BHM, MAB, MAD and CFR designed the study.

Conflicts of interest: SFS, MLW, BSD, MAB, MAD, SJA, FEL have no conflicts of interest to declare. Dr. Lenze reports current grants from NIH (NIA, NCCIH, NIMH, OBSSR), grants from FDA, grants from McKnight Brain Research Foundation, grants from Taylor Family Institute for Innovative Psychiatric Research, grants from Barnes Jewish Foundation, grants from Takeda, grants from Lundbeck, and past grants from Roche, Sidney R. Baer Foundation, that did not influence the submitted work. Dr. Reynolds III, MD reports grants from National Institute of Health (NIH), grants from Center for Medicare and Medicaid Services (CMS), grants from Patient Centered Outcomes Research Institute (PCORI), grants from John A. Hartford Foundation, grants from American Foundation for Suicide Prevention, grants from Commonwealth of Pennsylvania, grants from Clinical and Translational Science Institute (CTSI), grants from National Palliative Care Research Center (NPCRC), during the conduct of the study; personal fees from American Association for Geriatric Psychiatry, personal fees from UPMC Endowment in Geriatric Psychiatry, an speakers honorarium from Medscape/WebMD did not influence the submitted work; in addition, Dr. Reynolds III, MD has a patent Psychometric analysis of the Pittsburgh Sleep Quality Index (PSQI) with royalties paid to Dr. Daniel Buysse. Bristol Meyers Squibb, Forrest Labs, Lilly, Pfizer provide pharmaceutical supplies for NIH-sponsored work (these pharmaceutical companies play no role in the design, analysis, or reporting of my data in peer reviewed journals). Dr. Blumberger reports other from Tonika Magventure, grants from Canadian Institutes of Health Research, grants from National Institute of Health, other from Brain Research and Development Services, grants from Brain Canada, did not influence the submitted work. Dr. Mulsant reports grants from National Institute of Mental Health (NIMH, USA), non-financial support from Pfizer, non-financial support from Bristol-Myers-Squibb, during the conduct of the study; grants from Brain Canada, grants from Canadian Institutes of Health Research, grants from US National Institute of Health, grants from Centre for Addiction and Mental Health Foundation, non-financial support from Eli Lilly, non-financial support from General Electric, did not influence the submitted work. Dr. Karp reports receipt of medication supplies for investigator initiated studies from Pfizer and Invidior that did not influence the submitted work.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

⁶Department of Psychiatry and Behavioral Sciences, University of Texas Health Science Center at Houston, Houston, TX, USA

⁷Departments of Psychology, Epidemiology, Biostatistics, and Clinical and Translational Science, University of Pittsburgh, Pittsburgh, PA, USA

⁸Department of Behavioral and Community Health Sciences, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania

Abstract

Personalizing treatment for late-life depression requires identifying and integrating information from multiple factors that influence treatment efficacy (moderators). We performed exploratory moderator analyses using data from a multi-site, randomized, placebo-controlled, double-blind trial of aripiprazole augmentation. Patients (n=159) aged 60 years had major depressive disorder that failed to remit with venlafaxine monotherapy. We examined effect sizes of 39 potential moderators of aripiprazole (vs. placebo) augmentation efficacy using the outcome of percentage reduction in depressive symptom after 12 weeks. We then incorporated information from the individually relevant variables in combined moderators. A larger aripiprazole treatment effect was related to: white race, better physical function, better performance on Trail-Making, attention, immediate, and delayed memory tests, greater psychomotor agitation and suicidality symptoms, and a history of adequate antidepressant pharmacotherapy. A smaller aripiprazole treatment effect was observed in patients with: more pain and more work/activity impairment and libido symptoms. Combining information from race and Trail-Making test performance (base combined moderator (M_b^*)) produced a larger effect size (Spearman effect size=0.29 (95% confidence interval (CI): 0.15, 0.42)) than any individual moderator. Adding other individually relevant moderators in the full combined moderator (M_f^*) further improved effect size (Spearman effect size=0.39 (95% CI: 0.25, 0.52)) and identified a sub-group benefiting more from placebo plus continuation venlafaxine monotherapy than adjunctive aripiprazole. Combining moderators can help clinicians personalize depression treatment. We found the majority of our patients benefited from adjunctive aripiprazole, but a smaller subgroup that is identifiable using clinical measures appeared to benefit more from continuation venlafaxine plus placebo.

Keywords

Aripiprazole; late-life depression; moderators; combined moderators; personalized medicine

First-line pharmacological treatments fail for over 50% of older adults with major depressive disorder (Joel et al., 2014, Mulsant et al., 2014, Smagula et al., 2015). Safe and efficacious adjunctive treatments for late-life major depressive disorder (LLD) are in high demand for the rapidly aging global population (Cauley, 2013). However, there remains a dearth of information regarding which second-line treatments are efficacious and for whom.

Our group recently found aripiprazole augmentation is both safe and efficacious for older adults with major depressive disorder who failed to respond to 12-weeks of open-label venlafaxine XR treatment (Lenze et al., 2015). In subsequent analyses specified *a priori* in the ClinicalTrials.gov registration, we examined a limited number of factors that we

hypothesized might moderate aripiprazole efficacy (Kaneriya et al., 2016). We found set-shifting performance, a cognitive performance measure that reflects executive function that can be assessed with the Trail-Making test, moderated aripiprazole response, while response inhibition, anxiety severity, and medical co-morbidity did not. Among patients with intact set-shifting performance, adding aripiprazole to venlafaxine was associated with over four times the odds of remission (compared with venlafaxine plus placebo). In contrast, in the presence of set-shifting impairment, there was no significant benefit of adding (to venlafaxine) aripiprazole compared with placebo.

Because our prior report only examined four pre-specified factors (among which only set-shifting moderated aripiprazole response), our clinical characterization of patients who may benefit from aripiprazole remains limited. Furthermore, set-shifting alone is an imperfect predictor of aripiprazole efficacy: even among patients without set-shifting impairment, remission rates with aripiprazole remained relatively low (50.7%). In patients with set-shifting impairment, adjunctive aripiprazole was not statistically different from venlafaxine plus placebo. Because individual moderator effect sizes tend to be small and may provide conflicting treatment indications for a single individual, matching treatments to patients requires identifying multiple moderating variables and combining their information into a single, clinically useful index (of whether, for a given patient, which treatment is expected to be more efficacious than an alternative).

Recently, methods have been developed specifically for this task, allowing the integration of information from multiple moderators into single combined moderator with a stronger effect size than any of its individual components (Kraemer, 2013, Wallace et al., 2013). In the current work, we applied combined moderator methods in exploratory analyses of potential clinical moderators of aripiprazole's efficacy. We used clinical experience and past literature to select a pool of baseline characteristics that may moderate the effect of venlafaxine plus adjunctive aripiprazole versus venlafaxine plus placebo. We examined the roles of pain, physical function difficulties, and suicidality, which may mark etiopathological differences relevant to aripiprazole response. We also considered clinical characteristics previously related to first-line treatment outcomes (e.g., duration of the current depressive episode (Joel, Begley, 2014, Nelson et al., 2012, Smagula, Butters, 2015)). Variability in individual depressive symptom expression (i.e. items from the Hamilton Depression Rating Scale (Hamilton, 1960)) is associated with first-line treatment outcomes (Buhler et al., 2014, Smagula, Butters, 2015), but has yet to be examined as potential moderators of second-line treatments like aripiprazole. A primary side effect of aripiprazole includes akathisia, and we tested whether extrapyramidal symptoms (including akathisia) prior to treatment decreased the antidepressant efficacy of aripiprazole (e.g. due to an inability to tolerate a therapeutic dose). Finally, we expanded our search for the neuropsychological moderators of aripiprazole efficacy by examining, in addition to those pre-specified and previously examined (set-shifting and response inhibition), immediate memory, delayed memory, attention, language, and visuospatial performance. Our aims were to provide a comprehensive description of the clinical characteristics associated with aripiprazole efficacy, and to compute a combined moderator that achieved a greater effect size and enhanced discriminative utility compared with individual moderators alone.

Materials and Methods

Participants

Participants were aged 18–60 with a current major depressive episode (diagnosed by SCID/DSM-IV criteria) and a Montgomery Asberg Depression Rating Scale (MADRS) (Montgomery and Asberg, 1979) score of ≥ 15 . Individuals with a lifetime diagnosis of bipolar disorder, dementia, schizophrenia, and alcohol or substance abuse during the last 6 months, or current psychotic symptoms, were excluded from the study. A complete description of participants, including recruitment, has been published previously (Lenze, Mulsant, 2015). The conduct of the study was overseen by a Data Safety and Monitoring Board and all participants provided written informed consent.

Interventions

To prospectively establish treatment resistance, patients were first treated openly with venlafaxine XR (up to 300 mg/day) for 12 weeks. The current study pertains to the 181 patients who failed to achieve remission (defined by a MADRS score ≥ 10 for two sequential assessments) by the end of the open-label venlafaxine trial. Patients were then randomly assigned, using permuted block randomization, to the addition of aripiprazole or placebo for 12 weeks while maintaining the dose of venlafaxine achieved during initial monotherapy. The randomized augmentation phase was conducted under double-blind conditions with outcomes assessed by independent evaluators. No member of the research team other than the pharmacist was aware of treatment assignment. Aripiprazole or placebo tablets were started at 2 mg daily and titrated as tolerated, to a maximal dose of 15 mg daily.

Assessments

Outcome—We quantified reductions in depressive symptoms as the percentage change in depressive symptoms, measured with the MADRS before and after 12 weeks of aripiprazole or placebo augmentation. This continuous outcome provides increased statistical power when compared with categorical definitions of treatment response, and the methods to develop a combined moderator as described by Kraemer (Kraemer, 2013) require a continuous outcome. After computing combined moderators (see below), we compared their discriminative utilities using a clinically relevant and interpretable outcome of categorically defined treatment response (defined as achieving at least a 50% reduction in depressive symptoms by week 12 measured with the MADRS).

Potential moderators—All moderators were assessed before the randomized trial. Pain was measured with the Numeric Rating Scale for Pain (Herr and Mobily, 1993), physical function was measured with the physical function subscale of the Medical Outcomes Study 36-item (MOS-36) Health Survey (Ware and Sherbourne, 1992), and suicidality was measured with the Scale for Suicidal Ideation (SSI) (Beck et al., 1979). We examined the following depressive illness characteristics: age at onset of current episode, length of time since initial exposure to depressive illness (current age minus age at first episode), single vs. recurrent depression, and the duration of the current episode (in weeks). We also assessed the adequacy of past antidepressant treatment with the Anti-depressant Treatment History Form (ATHF; Oquendo et al. (2003); scores ≥ 3 indicate a prior adequate trial), which was

assessed prior to the open label trial where all participants were provided with an adequate trial of venlafaxine. Individual depressive symptoms were examined from the Hamilton Depression Rating Scale (Hamilton, 1960), which was also examined as a total score. We examined extrapyramidal symptoms using the Barnes Akathisia (Sachdev, 1994) and Simpson-Angus Scales (Knol et al., 2009). Baseline neuropsychological measures included domain-specific index scores (assessing visuospatial reasoning, immediate and delayed memory, attention, and language performance) from the Repeatable Battery for the Assessment of Neuropsychological Status Randolph et al. (1998) and two tasks from the Delis-Kaplan Executive Function System that tested set-shifting (Trail Making test condition 4 vs. 5 scaled scores) and response inhibition (Color-word interference condition 3) (Delis DC, 2001). We also assessed whether age, gender, education, or race (white vs. non-white) moderated aripiprazole efficacy. Complete covariate data were required for inclusion in the combined moderator analysis (n=22 were excluded).

Statistical analysis

We first computed individual Spearman correlation moderator effect sizes and 95% confidence intervals using bootstrapping following the method outlined by Kraemer (Kraemer, 2013) and previously implemented in a separate study (Wallace, Frank, 2013). Non-parametric Spearman correlation was used to allow for non-normal and dichotomous moderators, and to reduce the influence of potential outliers on the results. Because this is exploratory research, and treatment effect sizes are more relevant to clinical interpretations and decision-making than p-values alone, our focus is on effect sizes rather than p-values, thereby reducing type-1 error.

Next, we developed two combined moderators to evaluate whether computing a combined moderator from multiple clinical variables improved the effect size and discriminative utility over a basic combined moderator. At its core, a combined moderator is simply an optimally weighted combination of individual moderators. The base combined moderator (M_b^*) model included only the previously identified moderator (set-shifting performance measured with the Trail-Making test) plus easily obtainable sociodemographic characteristics. Next, in a full model (M_f^*), we entered all moderators with an individual Spearman moderator effect sizes ≥ 0.10 (i.e., at least a “small” effect size).

Preliminary methods to select which individual moderators to include in the combined moderator, as well as novel regression methods to obtain optimal weights that each individual moderator should contribute, have been developed and described previously (Kraemer, 2013). Here, we streamlined the process of variable selection and weight estimation through the use of Lasso regression (Tibshirani, 1994), which simultaneously performs variable selection and weight estimation by shrinking the weights corresponding to the least useful variables to zero. Ultimately, the individual moderators that are more useful moderators in the context of other moderators are given greater weights, and those that are not useful are removed from the model as a result of their weight of zero. Once weights are estimated, they are used to calculate a combined moderator score for each individual -- an optimally weighted linear combination of their individual moderator scores. For each combined moderator, we calculated the moderator effect size and 95% confidence interval.

We illustrate the observed effects by plotting moderators against the expected (predicted) proportion depressive symptom change stratified by treatment assignment.

Finally, we assessed the discriminative utility of these combined moderators using the c-statistic as estimated through a logistic regression of study site, treatment assignment, the combined moderator, and the combined moderator by treatment interaction on categorically defined treatment response (defined as 50% depressive symptom reduction). A c-statistic value of 0.5 represents chance discrimination, whereas 1.0 represents perfect discrimination.

Results

Patients were on average 67 years old (standard deviation, 6). Additional descriptive information was published previously (Kaneriyia, Robbins-Welty, 2016).

Individual moderator effects

Individual moderator effect sizes were small (Table 1) and included both positive and negative moderators. As values of positive moderators increase, the effect of aripiprazole (vs. placebo) augmentation increased. As we found previously, better performance on the Trail-Making test (set-shifting) was a positive moderator (i.e., related to greater aripiprazole efficacy (see Figure 1)); additionally, our results show that Trail-Making test performance was the largest individual moderator.

Other positive moderators were (Table 1): white (vs. non-white) race, attention, immediate, and delayed memory domain performance, having better physical function (i.e., higher MOS-36 scores on the physical subscale), higher psychomotor agitation (i.e., higher score on the HDRS item 9), higher suicidality (i.e., higher scores on the Suicidal Ideation Scale), and having received an adequate trial of antidepressant pharmacotherapy (ATHF 3; before the open-label trial). In patients with these characteristics we observed a greater benefit of aripiprazole over placebo.

We also identified several negative moderators: as values of the negative moderators increase, the effect of aripiprazole (vs. placebo), the benefit of aripiprazole was diminished. The strongest negative aripiprazole moderators included higher pain ratings, higher “work/activity impairment” (i.e., higher score on the HDRS item 7), and lower libido (i.e., higher score on the HDRS item 14).

Combined moderator effects

The base combined moderator (M_b^*) had a larger effect size than any individual moderator (M_b^* effect size=0.29 (95% CI: 0.15, 0.42)). This effect size was improved further in the full combined moderator (M_f^* effect size=0.39 (95% CI: 0.25, 0.53)). We provide weights (Supplemental Table) that can be used to calculate these combined moderators for any patient given measures of the relevant clinical variables.

Figure 1 illustrates the moderating effects of set-shifting and the two combined moderators; each plot shows the expected proportion depressive symptom change (y-axis) by different levels of the moderators (x-axis), separately for aripiprazole and placebo arms. For all three

of these moderators, higher moderator values were associated with a greater benefit of adjunctive aripiprazole over placebo; conversely, lower values tended to be associated with a greater benefit of placebo over adjunctive aripiprazole.

We categorized patients above and below the cross-point of the base combined moderator (M_b^*). Above the cross-point, aripiprazole was associated with greater symptom reductions compared with placebo (above M_b^* cross-point (n=119): aripiprazole vs. placebo effect=0.54, 95% CI: 0.20, 0.93). In patients below this cross-point (n=40), we found no statistically significant effect of treatment arm (aripiprazole treatment effect=-0.41, 95% CI: -1.09, 0.19).

Patients were also categorized based on whether they were above or below the cross-point of the full combined moderator (M_f^*). The treatment effect was reversed above and below this cross-point: in patients above this cross-point, aripiprazole was associated with a significantly greater proportion symptom reduction than placebo (n=111 who 'benefit from aripiprazole (plus venlafaxine)'; aripiprazole vs. placebo effect=0.73, 95% CI: 0.36, 1.17); in patients below this cross-point, aripiprazole was associated with a significantly smaller symptom reduction when compared with placebo (n=48 who 'better with placebo (plus venlafaxine)'; aripiprazole treatment effect=-0.66, 95% CI: -1.40, -0.07).

To illustrate the relevant clinical characteristics related to the efficacy of adjunctive aripiprazole, and because the full moderator (M_f^*) provided the largest effect size, we provide descriptive information (Table 2) regarding the individual moderators selected in patients above and below the point where the lines cross in Figure 1c. On average, patients below this crosspoint (who were 'better with placebo') were less often white, tended to have lower cognitive scores (across set-shifting, attention, and memory domains), more pain, lower physical function scores, greater "work/activity impairment," less agitation, less suicidal ideation, and had less often received a prior adequate trial of antidepressant pharmacotherapy.

Table 2 also includes other important clinical information. For example, among patients who were above the cut-point (i.e., those who were expected to "benefit from aripiprazole"), the number needed to treat with aripiprazole was 4.0 for both response and remission outcomes; patients below this cut-point had negative number needed to treat values, indicating aripiprazole was not superior over placebo in this sub-group. Table 2 also shows that patients above and below this cut-point did not differ in terms of the rate of treatment-emergent akathisia (the most common adverse effect we previously identified (Lenze, Mulsant, 2015)).

Comparison of moderators' discriminative utility

With categorically defined response as the outcome, a model with only the previously identified moderator (Trail-Making test performance) corresponded to a c-statistic of 0.70. Discriminative utility improved by adding demographic factors in the base combined moderators (M_b^* c-statistic=0.74) and was similar in the full combined moderator (M_f^* c-statistic=0.75).

Discussion

Our results provide a novel description of the distinct clinical profiles which characterize patients who are, and are not, expected to benefit from adjunctive aripiprazole pharmacotherapy for late-life major depression that does not respond to 12-weeks of venlafaxine treatment. We identified the clinical factors related the efficacy of aripiprazole, and provide a means (Supplemental Table) for combining the relevant information into a single indication of whether a given patient is likely to benefit from adding aripiprazole to venlafaxine. Combining information from these multiple individual moderators resulted in improved effect size and discriminative utility when compared with using the strongest individual moderator alone. We encourage clinicians and researchers to compute values of these combined moderators in their patients/samples to assess the generalizability and clinical utility of our findings, and move towards evidence-based decision practices for the prescription of second-line pharmacotherapy for late-life depression.

With categorically defined treatment response as the outcome, we observed similar c-statistics between the base (M_b^* , including Trail-Making test performance and socio-demographic characteristics) and full moderators (M_f^*). This suggests combining information about race and set-shifting may provide the most useful, easily obtainable indicator of whether a patient is likely to respond to adjunctive aripiprazole. Nevertheless, the full moderator (M_f^*) had a larger effect size compared with the base moderator (M_b^*) in terms of association with the continuous level of change in depression severity. Indeed, the full moderator (M_f^*) identified a sub-group wherein adding placebo to venlafaxine was associated with greater symptom reductions than when adding aripiprazole. In contrast, among patients by which Trail-Making performance alone or the base combined moderator (M_b^*) indicated against aripiprazole, there was no statistically significant difference between the treatment arms' outcomes.

The full combined moderator, M_f^* , identified the largest group with indication against aripiprazole's potential efficacy (n=48 below the M_f^* cross-point, compared with n=40 below the base combined moderator (M_b^*) cross-point). Thus, drawing on information from additional sources, the combined moderator selected 8 more patients from the majority in whom aripiprazole is, on average, efficacious. This same practice might prove useful clinically; for example, clinicians may be rightly cautious assuming aripiprazole's efficacy in patients with set-shifting impairment reflected on the Trail-Making test, *as well as* in patients with normative set-shifting performance accompanied by a high level of pain and work/activity impairment.

We observed moderating roles of set-shifting performance, attention, and memory function, which may suggest that a common neurobiological substrate is necessary for both these cognitive functions and aripiprazole's antidepressant effect. Set-shifting had the largest individual effect size, and because it is most easily measured, it may be most clinically useful. Nevertheless, our findings suggest depression accompanied by mild cognitive impairment (e.g., psychometrically defined as one standard deviation below normative performance, see Table 2) across multiple domains of cognitive function will not respond to aripiprazole; potentially, in these patients, aripiprazole cannot engage relevant neural

circuitry (e.g., dopamine receptors in frontal-striatal circuitry) and/or resolve the underlying pathology (e.g., neurodegenerative disease) that contributes to the generation and perpetuation of the depressive episode.

We found non-white race was associated with aripiprazole being less efficacious. A prior randomized controlled trial comparing aripiprazole and placebo for treatment resistant mid-life depression found no significant interaction between race and treatment assignment, but did report that among mid-life adults there was no statistically significant benefit of aripiprazole in non-whites sub-groups (Thase et al., 2008). This is consistent with our observation regarding the moderating role of non-white race. Future research is needed to better understand what accounts for the reduced efficacy we observed among non-white patients.

We also found patients with worse physical function, higher levels of pain, lower libido symptoms, and work/activity impairment were less likely to benefit from aripiprazole. The full combined moderator separated a sub-group, characterized by these factors, which benefited more from venlafaxine treatment plus placebo than with the addition of aripiprazole. Potentially, these factors mark etiological differences in depression that require treatments targeted to, or at least addressing, the underlying stressors that trigger or perpetuate the depressive episode (e.g., chronic pain). This suggests that patients with these characteristics may require an extended course (e.g., 24 weeks) of venlafaxine. In addition, patients with higher levels of agitation and suicidality appeared to benefit more from aripiprazole, suggesting aripiprazole may be particularly efficacious for treatment resistant depression with these symptoms.

Finally, having received an adequate trial of antidepressant pharmacotherapy before enrolling in our study was associated with greater efficacy of augmentation with aripiprazole compared with placebo. This finding is consistent with our recent report on this same trial (Hsu, in press), which found that when given continuation venlafaxine plus placebo, having a prior history of antidepressant treatment was associated with a lower likelihood of achieving remission; but when adding aripiprazole, having a prior history of antidepressant treatment was not related to differences in remission rates. This prior work did not statistically test whether a prior history of antidepressant treatment moderated the efficacy of aripiprazole vs. placebo. Our paper now adds that, indeed, having a prior history of antidepressant treatment moderates the efficacy of aripiprazole vs. placebo augmentation. Thus, for patients with a history of prior antidepressant treatment that do not respond to an initial trial of venlafaxine, adding a second-line treatment like aripiprazole may be necessary. Furthermore, the current work now places this information in the context of several other additional and important moderators.

Several limitations should be noted. Our findings must be validated in an independent sample. In addition, with an average age of about 67 years, our findings cannot be generalized to the “oldest-old.” Finally, future mechanistic research is needed to directly test and elucidate the pathways by which the observed moderating variable relate to aripiprazole efficacy.

In conclusion, our findings suggest set-shifting impairment and non-white race may be the most useful, easily obtainable clinical indicators that a patient is unlikely to benefit from adjunctive aripiprazole. Our results also suggest taking into account the other clinical moderators detected (including performance on other cognitive tests, physical function, pain, and individual symptoms) can help identify additional patients who are likely to benefit more from continued venlafaxine monotherapy than from the addition of aripiprazole.

Future research is needed to replicate these findings, implement replicated findings clinically, investigate the mechanisms of the observed treatment and moderation effects, and use randomized experiments to identify effective treatments for the subgroups of patients who are unlikely to respond to aripiprazole. If future external validation is successful, when treating patients with low M^*_F values (e.g. those who were ‘better with placebo (plus venlafaxine)’), clinicians would be advised continue closely monitored monotherapy, or to utilize a second-line treatment that targets the underlying source of mood disturbance reflected in the patient’s clinical profile.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was sponsored by the following National Institute of Mental Health grants: R01MH083660, R01MH083648, R01MH083643, P30MH90333, T32MH19986.

Role of Funding Source: The funding source had no role in the study design; the collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication.

References

- Beck AT, Kovacs M, Weissman A. Assessment of suicidal intention: the Scale for Suicide Ideation. *Journal of consulting and clinical psychology*. 1979; 47:343–52. [PubMed: 469082]
- Buhler J, Seemuller F, Lage D. The predictive power of subgroups: an empirical approach to identify depressive symptom patterns that predict response to treatment. *Journal of affective disorders*. 2014; 163:81–7. [PubMed: 24836092]
- Cauley, J. The Demography of Aging. In: Newman, ACJ., editor. *The Epidemiology of Aging*. Springer; 2013. p. 3–14.
- Delis, DCKE.; Kramer, JH. *Dellis Kaplan Executive Function System Examiner’s Manual*. The Psychological Corporation; 2001.
- Hamilton M. A rating scale for depression. *Journal of Neurology, Neurosurgery and Psychiatry*. 1960; 23:56–62.
- Herr KA, Mobily PR. Comparison of selected pain assessment tools for use with the elderly. *Applied nursing research : ANR*. 1993; 6:39–46. [PubMed: 8439177]
- Hsu J, Mulsant, Lenze, Karp, Lavretsky, Roose, Reynolds, Blumberger. Impact of Prior Treatment on Remission of Late-Life Depression with Venlafaxine and Subsequent Aripiprazole or Placebo Augmentation. *American Journal of Geriatric Psychiatry*. in press.
- Joel I, Begley AE, Mulsant BH, Lenze EJ, Mazumdar S, Dew MA, et al. Dynamic Prediction of Treatment Response in Late-Life Depression. *The American journal of geriatric psychiatry : official journal of the American Association for Geriatric Psychiatry*. 2014; 22:167–76. [PubMed: 23567441]

- Kaneriya SH, Robbins-Welty GA, Smagula SF, Karp JF, Butters MA, Lenze EJ, et al. Predictors and Moderators of Remission With Aripiprazole Augmentation in Treatment-Resistant Late-Life Depression: An Analysis of the IRL-GREY Randomized Clinical Trial. *JAMA psychiatry* (Chicago, Ill). 2016; 73:329–36.
- Knol W, Keijsers CJ, Jansen PA, Belitser SV, Schobben AF, Egberts AC, et al. Validity and reliability of the Simpson-Angus Scale (SAS) in drug induced parkinsonism in the elderly. *International journal of geriatric psychiatry*. 2009; 24:183–9. [PubMed: 18615779]
- Kraemer HC. Discovering, comparing, and combining moderators of treatment on outcome after randomized clinical trials: A parametric approach. *Statistics in medicine*. 2013; 32:1964–73. [PubMed: 23303653]
- Lenze E, Mulsant BH, Blumberger DM, Karp JF, Newcomer JW, Anderson SJ, et al. Efficacy, safety, and tolerability of augmentation pharmacotherapy with aripiprazole for treatment-resistant depression in late life: a randomized placebo-controlled trial. *Lancet*. 2015
- Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *The British journal of psychiatry : the journal of mental science*. 1979; 134:382–9. [PubMed: 444788]
- Mulsant BH, Blumberger DM, Ismail Z, Rabheru K, Rapoport MJ. A systematic approach to pharmacotherapy for geriatric major depression. *Clinics in geriatric medicine*. 2014; 30:517–34. [PubMed: 25037293]
- Nelson JC, Zhang Q, Deberdt W, Marangell LB, Karamustafalioglu O, Lipkovich IA. Predictors of remission with placebo using an integrated study database from patients with major depressive disorder. *Current medical research and opinion*. 2012; 28:325–34. [PubMed: 22292447]
- Oquendo MA, Baca-Garcia E, Kartachov A, Khait V, Campbell CE, Richards M, et al. A computer algorithm for calculating the adequacy of antidepressant treatment in unipolar and bipolar depression. *The Journal of clinical psychiatry*. 2003; 64:825–33. [PubMed: 12934985]
- Randolph C, Tierney MC, Mohr E, Chase TN. The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS): preliminary clinical validity. *Journal of clinical and experimental neuropsychology*. 1998; 20:310–9. [PubMed: 9845158]
- Sachdev P. A rating scale for acute drug-induced akathisia: development, reliability, and validity. *Biological psychiatry*. 1994; 35:263–71. [PubMed: 7910487]
- Smagula SF, Butters MA, Anderson SJ, Lenze EJ, Dew MA, Mulsant BH, et al. Antidepressant Response Trajectories and Associated Clinical Prognostic Factors Among Older Adults. *JAMA psychiatry* (Chicago, Ill). 2015
- Thase ME, Trivedi MH, Nelson JC, Fava M, Swanink R, Tran Q-V, et al. Examining the Efficacy of Adjunctive Aripiprazole in Major Depressive Disorder: A Pooled Analysis of 2 Studies. *Primary Care Companion to The Journal of Clinical Psychiatry*. 2008; 10:440–7.
- Tibshirani R. Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society, Series B*. 1994:267–88.
- Wallace ML, Frank E, Kraemer HC. A novel approach for developing and interpreting treatment moderator profiles in randomized clinical trials. *JAMA psychiatry* (Chicago, Ill). 2013; 70:1241–7.
- Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Medical care*. 1992; 30:473–83. [PubMed: 1593914]

Figure 1.

Illustration of interactions between three moderator variables with treatment assignment in relation to the predicted change in depression symptoms over the 12-week randomized controlled trial of adjunctive aripiprazole (ARI) vs placebo (PBO)

Table 1

Individual moderator effect sizes and 95% bootstrap CIs

Moderator	Spearman effect size (95% CI)
Trail-Making test (condition 4 vs. 5 scaled scores) *	0.196 (0.060, 0.330)
White (vs. non-white) *	0.194 (0.070, 0.320)
RBANS Attention domain *	0.166 (0.010, 0.309)
RBANS Immediate memory domain *	0.153 (0.011, 0.290)
Numeric Rating Scale for Pain *	-0.153 (-0.290, 0.008)
MOS - Physical component score *	0.121 (-0.040, 0.261)
“Work/activity impairment” (HDRS item 7) *	-0.109 (-0.255, 0.044)
RBANS Delayed memory domain *	0.108 (-0.033, 0.245)
“Agitation” (HDRS item 9) *	0.104 (-0.043, 0.247)
Scale for Suicidal Ideation *	0.099 (-0.020, 0.218)
“Genital symptoms” (HDRS item 14) *	-0.097 (-0.234, 0.050)
Past adequate antidepressant treatment (ATHF 3)	0.095 (-0.066, 0.249)
“Psychomotor retardation” (HDRS item 8)	-0.092 (-0.237, 0.044)
Education (in years)	0.083 (-0.058, 0.215)
Age	0.082 (-0.085, 0.229)
Simpson-Angus Scale at randomization	-0.082 (-0.227, 0.062)
Length of time since first exposure to depression (years from first episode)	0.081 (-0.075, 0.223)
“Feelings of guilt” (HDRS item 2)	0.071 (-0.074, 0.207)
“Depressed mood” (HDRS item 1)	-0.070 (-0.219, 0.069)
“Insomnia (late)” (HDRS item 6)	0.064 (-0.074, 0.204)
“Anxiety (somatic)” (HDRS item 11)	0.062 (-0.078, 0.193)
Age at first episode	-0.060 (-0.210, 0.099)
“Suicide” (HDRS item #3)	0.058 (-0.067, 0.181)
“Somatic symptoms (gastrointestinal)” (HDRS item 12)	-0.057 (-0.182, 0.080)
Color-word interference (condition 3)	0.054 (-0.095, 0.193)
Male (vs. female) sex	-0.048 (-0.192, 0.101)
Episode duration (in weeks)	-0.047 (-0.192, 0.077)
“Insomnia (early)” (HDRS item 4)	0.042 (-0.099, 0.174)
“Anxiety (psychological)” (HDRS item 10)	0.034 (-0.101, 0.167)
RBANS Language domain	0.029 (-0.129, 0.179)
“Insomnia (middle)” (HDRS item 5)	0.025 (-0.124, 0.183)
“Loss of weight” (HDRS item 16)	0.023 (-0.159, 0.198)
“Insight” (HDRS item 17)	-0.019 (-0.142, 0.099)
Recurrent vs. single episode	-0.010 (-0.156, 0.141)
Barnes Akathisia Scale	0.010 (-0.161, 0.133)
RBANS Visuospatial domain	0.007 (-0.153, 0.171)
“Somatic symptoms general” (HDRS item 13)	-0.006 (-0.151, 0.155)

Moderator	Spearman effect size (95% CI)
"Hypochondriasis" (HDRS item 15)	0.006 (−0.124, 0.141)
HDRS (total score)	0.004 (−0.133, 0.151)

Positive moderator effect sizes indicate the effect of aripiprazole is higher among patients with greater values on that variable (e.g., greater performance on the Trail-Making test were related to a greater effect of aripiprazole on the percent symptom reduction);

* indicates variable was selected into the full combined moderator;

Table 2

Descriptive characteristics of patients expected to respond ‘better with placebo’ or ‘benefit from Aripiprazole’ when added to venlafaxine (i.e. those below and above $M_F^*=0.0002$)

	‘Better with placebo (plus venlafaxine)’ ($M^*<0.0002$; n=48)	‘Benefit from aripiprazole (plus venlafaxine)’ ($M^*>0.0002$; n=111)
Selected moderators		
White race, % (n)	73 (35)	96 (107)
Past adequate antidepressant treatment (ATHF 3)	63 (30)	78 (87)
Trail making test (condition 4 vs. 5 scaled scores)	5.73 (3.19)	9.86 (2.9)
RBANS Attention domain	90.02 (16.73)	103.99 (16.52)
RBANS Immediate memory domain	86.44 (18.84)	101.32 (14.5)
RBANS Delayed memory domain	86.44 (16.06)	101.2 (12.65)
Numeric Rating Scale for Pain	10.08 (4.6)	5.14 (4.74)
MOS - Physical component	37.73 (10.84)	44.77 (12.18)
Suicidal Ideation Scale	0.60 (1.5)	2.00 (4.75)
“Work/activity impairment” (HDRS item 7)	2.71 (0.85)	2.37 (0.94)
“Agitation” (HDRS item 9)	0.06 (0.24)	0.32 (0.54)
“Genital symptoms” (HDRS item 14)	0.85 (0.9)	0.41 (0.73)
Other clinical information		
Response rate given aripiprazole, % (n)	12 (25)	59 (34)
Response rate given placebo, % (n)	48 (11)	26 (14)
Number needed to treat (response)	−3.0	4.0
Remission rate given aripiprazole, % (n)	16 (4)	57 (33)
Remission rate given placebo, % (n)	39 (9)	26 (14)
Number needed to treat (remission)	−5.0	4.0
At least a two point increase in the UKU akathisia item during treatment, % (n)	8.33 (4)	10.81 (12)

Mean (SD) unless otherwise noted; Remission defined as achieving at least MADRS 10 for at least the final two visits of the trial;