ORIGINAL RESEARCH

Venlafaxine XR treatment for older patients with major depressive disorder: decision trees for when to change treatment

Helena Kyunghee Kim ,[1] Daniel M Blumberger,[1,2] Jordan F Karp,[3] Eric Lenze,[4] Charles F Reynolds,[5] Benoit H Mulsant[1,2]

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/ebmental-2022-300479).

[1]Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada
[2]Centre for Addiction and Mental Health, Toronto, Ontario, Canada
[3]Department of Psychiatry, University of Arizona, Tucson, Arizona, USA
[4]Department of Psychiatry, University of Washington, St. Louis, Missouri, USA
[5]Department of Psychiatry, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

Correspondence to
Professor Benoit H Mulsant, Psychiatry, University of Toronto, Toronto, ON M5T 1R8, Canada; benoit.mulsant@utoronto.ca

Received 23 March 2022
Accepted 31 August 2022
Published Online First 13 September 2022

To cite: Kim HK, Blumberger DM, Karp JF, *et al*. *Evid Based Ment Health* 2022;**25**:156–162.

ABSTRACT

Background Predictors of antidepressant response in older patients with major depressive disorder (MDD) need to be confirmed before they can guide treatment.

Objective To create decision trees for early identification of older patients with MDD who are unlikely to respond to 12 weeks of antidepressant treatment, we analysed data from 454 older participants treated with venlafaxine XR (150–300 mg/day) for up to 12 weeks in the Incomplete Response in Late-Life Depression: Getting to Remission study.

Methods We selected the earliest decision point when we could detect participants who had not yet responded (defined as >50% symptom improvement) but would do so after 12 weeks of treatment. Using receiver operating characteristic models, we created two decision trees to minimise either false identification of future responders (false positives) or false identification of future non-responders (false negatives). These decision trees integrated baseline characteristics and treatment response at the early decision point as predictors.

Finding We selected week 4 as the optimal early decision point. Both decision trees shared minimal symptom reduction at week 4, longer episode duration and not having responded to an antidepressant previously as predictors of non-response. Test negative predictive values of the leftmost terminal node of the two trees were 77.4% and 76.6%, respectively.

Conclusion Our decision trees have the potential to guide treatment in older patients with MDD but they require to be validated in other larger samples.

Clinical implications Once confirmed, our findings may be used to guide changes in antidepressant treatment in older patients with poor early response.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Several studies have shown that clinical characteristics, such as early treatment response and frailty, can be used to predict antidepressant treatment response in older patients with major depressive disorder (MDD). However, when and how to use these predictors to guide treatment decisions need to be operationalised.

WHAT THIS STUDY ADDS

⇒ Our findings confirm that week 4 is the optimal time to consider treatment changes for older adults with MDD who have not yet achieved treatment response. Our analyses show the potential benefit of using decision trees combining baseline clinical factors and early change in depressive symptoms to guide clinicians about whether these older depressed patients should stay on the same antidepressant or a treatment change should be considered.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Our findings show the potential of decision trees to decide which older patients could benefit from early antidepressant treatment adjustments. Future studies validating our findings and clinical trials assessing the impact of using these prediction models are needed before they can be applied to clinical practice.

BACKGROUND

Most patients with major depressive disorder (MDD) do not respond to the first prescribed medication; therefore, a key approach to successful treatment is to advance patients along a treatment algorithm, switching or augmenting when the current medication prescribed at adequate dosage and duration is not relieving symptoms. Measurement-based care (MBC) using regular monitoring with validated scales can facilitate these treatment changes based on symptomatic change.[1] There is a growing body of studies describing how we can use this type of data to predict treatment response and minimise time receiving an ineffective antidepressant.[1–3] One approach involves creating a decision tree that can be employed early in course of treatment to identify patients who are unlikely to respond to their current treatment, and therefore, should receive a different treatment. Finding the optimal time point to employ such a model requires balancing minimising prolonged suffering on a potentially unhelpful antidepressant against preventing premature abandonment of an antidepressant that would have been effective with longer treatment.[4]

Greater comorbid physical burden,[5 6] comorbid anxiety[7 8] and recurrent depressive episodes[8 9] have been identified as predictors of treatment resistance in older adults with MDD. Several previous studies

have shown that these predictors can be combined with early symptomatic improvement to estimate the likelihood of eventual treatment response in these patients.[4 10 11]

OBJECTIVE

The aim of this analysis was to replicate and expand on these prior findings and construct decision trees that can identify early in the course of treatment older patients who are unlikely to attain treatment response with continued antidepressant treatment. To do this, we examined older patients receiving up to 12 weeks of treatment with venlafaxine extended release (XR) in the Incomplete Response in Late-Life Depression: Getting to Remission (IRL-GRey) study (ClinicalTrials.gov identifier NCT00892047).[12] The IRL-GRey dataset was used due to its relatively large sample size, frequent assessments and standardised approach to dosage titration. First, we aimed to identify the earliest decision point when non-responders and partial responders could be reliably detected. Then, we created decision trees for non-responders and partial responders at the early decision point. Finally, we constructed decision trees with baseline demographic and clinical factors as potential predictors of eventual treatment response (ie, at week 12).

METHODS

Overview

Details on methods and participant characteristics of IRL-GRey have been published elsewhere.[11 12] Briefly, the IRL-GRey dataset included 454 patients 60 years and older with MDD according to the criteria of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)[13] confirmed with the Structured Clinical Interview for DSM-IV[14] and at least moderate depressive symptoms as reflected by a Montgomery-Asberg Depression Rating Scale (MADRS)[15] score ≥15. Patients were excluded if they had dementia, bipolar disorder, schizophrenia, current psychotic symptoms or substance abuse or dependence within the past 6 months.

Participants were assessed with the MADRS at weeks 1, 2, 4, 6, 8, 10 and 12. They were treated with venlafaxine XR starting at 37.5 mg/day, titrated to 150 mg/day over 2 weeks as tolerated; if remission was not attained after 6 weeks, venlafaxine XR was further increased up to 300 mg/day for up to six additional weeks. Participants completed open treatment with venlafaxine XR either when they attained remission (defined as a MADRS score ≤10 for two consecutive visits) or when they completed 12 weeks of treatment.

Participant characteristics

We included in our analysis the following participant characteristics obtained at baseline as potential predictors of treatment response: age; self-reported sex at birth; self-reported race (white vs other); years of education; burden of comorbid physical illness measured with the Cumulative Illness Rating Scale-Geriatrics (CIRS-G)[16]; diagnosis of hypertension, heart disease or diabetes; severity of depressive symptoms measured with the MADRS total score; suicidality defined as a score ≥3 on the MADRS suicide item; severity of comorbid anxiety measured with the Brief Symptom Inventory (BSI)[17] anxiety score; not having responded to at least one previous adequate antidepressant trial before starting venlafaxine XR (as opposed to being treatment naïve or having only inadequate antidepressant trials) based on a score of 3 or higher on the antidepressant treatment history form[2 18]; duration of current episode; single or recurrent depression status; and age at onset of first MDD episode. These variables were chosen based on previous studies showing their association with antidepressant outcomes in the IRL-GRey dataset[2 11] and in other treatment studies of MDD in younger[5 8] or older adults.[2 5 7 11 18 19]

For this analysis, the primary outcome was a full antidepressant response, defined as a decrease in MADRS score higher than 50% from baseline. This outcome and this definition were used as they have been shown to be suitable for providing MBC and they do not depend on the scale used.[3] Based on previous studies,[4 20] partial response was defined as a decrease in MADRS score of 25%–50% from baseline and non-response as a decrease in MADRS score of less than 25%.

Data analysis

All analyses (except for the receiver operating characteristic (ROC) analysis) were performed using IBM SPSS Statistics 26. Because length of treatment was variable (ie, some participants dropped out before they completed the study, while other participants completed the study when they attained remission), imputations were used for both intermittent missing and monotone missing observations using the Markov Chain Monte Carlo option of SPSS multiple imputation procedure. Missing values were replaced with an average of five imputations, and a full dataset was created for the 454 participants as we have done in previous analyses.[4 10 11] Imputation details, including t-tests comparing original and imputed data, can be found in online supplemental table 1).

Identification of an early decision point

This first analysis aimed at identifying a time point that would fulfil the following three conditions: (1) more than 40% of partial responders at this time point attain full response at the end of treatment (ie, there is still hope for partial responders); (2) less than 25% of non-responders at this time point attain full response at the end of treatment (ie, most non-responders have been identified); (3) the proportion of full-responders at this time point with full response status at week 12 is sustained (ie, full-response is sustained). We call this time point the 'early decision point'. Condition #1 used 40% as a threshold because placebo-controlled clinical trials typically report antidepressant response rates of about 40% in geriatric depression.[21] Condition #2 was added because a recent meta-analysis showed that a third of patients with MDD and early non-improvement fully respond when treatment is extended to 12 weeks.[22] Thus, we set the threshold to a quarter of participants for criterion 2, which is a smaller minority. Condition #3 was added because some patients showing very early response revert to non-response status with longer treatment as their initial improvement may represent a 'placebo effect' rather than a true antidepressant effect.[23]

To identify an early decision point, for each assessment point up to week 10 (ie, weeks 1, 2, 4, 6, 8, 10), participants were divided into three groups: full responders, partial responders and non-responders. Then, for condition #1, we examined the proportion of partial responders at each assessment point who attain full response at week 12; and for condition #3, we examined the proportion of full-responders at each assessment point who maintain full response at week 12. For condition #2, as in our previous work,[4] we divided partial responders and non-responders at each assessment point up to week 10 (weeks 1, 2, 4, 6, 8, 10) and, for each of these assessment points, we calculated the proportion of participants who attain full response after various additional lengths of treatment up to week 12. For example, for participants at week 1, they will have additional

lengths of treatment of 1, 3, 5, 7, 9 and 11 weeks. For participants at week 8, they will have additional lengths of treatment of 2 and 4 weeks, adding up to 12 weeks in total. Finally, for each length of treatment, we calculated weighted mean proportion of participants who attained full response, with number of participants within each group (partial vs non-responder) at each assessment point used as weights.

Generalised estimating equations for comparing the effect of added lengths of treatment in partial responders and non-responders at the early decision point

After selecting the early decision point that met all three conditions, repeated measures model with generalised estimating equation (GEE) was performed to statistically compare the proportions of partial responders and non-responders at this early decision point who attained full response with each added length of treatment.[4] We used the GEE approach because GEE, which is a population-average model, accounts for within-group non-independence of observations and estimates the average response of the population within a group. As we were interested in estimating group effects, rather than model subject-specific effects, GEE was deemed appropriate.[24] Group (partial responders vs non-responders at the early decision point) and time effects (additional lengths of treatment in weeks) on attainment of full response were examined. The same comparison was repeated for each assessment point after the early decision point to determine whether added lengths of treatment would have a different effect in differentiating partial responders and non-responders at later assessment points.

Decision trees for predicting treatment response after 12 weeks of treatment at the early decision point

In a previous analysis, we identified demographic and clinical predictors of remission in all IRL-GRey participants, using a priori the change in MADRS after 2 weeks of treatment as one of the potential predictors.[11] Having at least one previous adequate antidepressant trial, baseline MADRS score, and improvement in MADRS score after 2 weeks were identified as significant predictors.[11] Our analysis expanded on these findings by identifying predictors of treatment response in participants who did not attain full response (ie, partial and non-responders) by the early decision point. We did this because a clinician would not change antidepressant treatment in a patient who has already attained full treatment response. For this analysis, we considered all the baseline patient characteristics discussed above plus response status (partial or non-response) at the early decision point. The Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis checklist[25] for reporting multivariate models can be found in online supplemental material 1.

We judged that our sample size was adequate for examining 14 predictors based on the rule of thumb of 10 events per variable (EPV).[26] Sample size calculation was also performed using a method proposed by Riley *et al* for binary outcomes (ie, logistic regression)[27] post hoc. Details of this sample size calculation can be found in online supplemental material 2.

We used ROC V.5.07 (https://web.stanford.edu/~yesavage/ROC.html) with an ROC curve analysis based on a modelling strategy where the programme searches all the predictor variables and identifies the optimal predictor variable that best predicts the outcome of interest using signal detection theory, weighting the importance of sensitivity and specificity.[28] Predictors were placed in a decision tree where the highest predicting variable divides the sample into two subsamples, and the process is repeated until the lowest predicting variable is found using a stopping rule of $p<0.05$. This approach is useful for analyses where predictors are likely to have high collinearity.[28] We developed two decision trees using sensitivity cutoffs of 0.3 ('low') and 0.7 ('high'). A low sensitivity decision tree minimises false positives (ie, falsely identifying a participant as someone who would attain eventual treatment response when they would not). This type of tree should be used for patients who are at high risk, such as inpatients or patients at risk of suicide. A high sensitivity decision tree minimises false negatives (ie, falsely identifying a participant as someone who would not attain eventual treatment response when they would). This type of tree should be used for patients in whom clinicians want to minimise medication changes, such as outpatients who have had multiple unsuccessful trials or frail outpatients.[10] Negative predictive value (NPV, ie, ability to predict eventual non-response) of the leftmost terminal node, representing the NPV of the combination of predictors identifying the subgroup of participants who are most likely to not reach eventual treatment response, is presented as a measure of model performance. NPV, positive predictive value (PPV) and accuracy of the overall decision tree are presented for general information. Decision trees trained on the complete dataset are presented. A fivefold cross-validation was performed to test the performance of the two decision trees. We report the average of the five test NPVs of the leftmost terminal nodes; the average of the five test NPVs of the overall trees; test PPVs of the overall trees; test accuracies of the overall trees; and the predictors of the five training decision trees for low and high sensitivity cutoffs.

FINDINGS

Participant characteristics

Of 454 participants, 197 participants (43.4%) attained full treatment response at week 12. Table 1 presents the baseline demographic and clinical characteristics of all participants and compares the full responders versus partial and non-responders at week 12. Full responders were more likely to be females; not to have a previous adequate antidepressant trial and have shorter duration of their current episode.

Selection of an early decision point

Looking at partial responders in table 2, more than 40% of partial responders at weeks 1–6 and attained full response at week 12; so according to our first criterion, the early decision point could be as early as week 1 and as late as week 6. Looking at non-responders in table 2, less than 25% non-responders at week 4 and in subsequent assessment points attained full response at week 12; so according to our second criterion, the early decision point could be as early as week 4. Finally, looking at full responders, 67.1% of full responders at week 2 maintained full response at week 12 and this proportion increased to 75.2% at week 4; after week 4, the proportion remained around 75%; so according to our third criterion, the early decision point could be as early as week 4. Therefore, week 4 was selected as the optimal early decision point.

Also, as shown in table 3B, more than a third of non-responders at weeks 1 and 2 attained full response with additional lengths of treatment. By contrast, less than a quarter of non-responders at week 4 attained full response with additional lengths of treatment, confirming the suitability of week 4 as our early decision point (ie, a point when 'hope' remains also for non-responders).

GEE comparing the effect of additional lengths of treatment in partial responders and non-responders at the early decision point

At the early decision point (ie, week 4), of 454 participants, 121 were full responders, 131 partial responders and 202

Table 1 Baseline demographic and clinical characteristics of all participants and of those who attained or did not attain full response at week 12*

	All participants	Full responders	Partial and non-responders	Statistics, p value†
N (%)	454 (100)	197 (43.4)	257 (56.6)	
Age (years)	69.0 (7.2)	69.5 (7.1)	68.7 (7.3)	U=27 393.5, 0.133
Self-reported sex (% female)	65.4 (N=297)	70.6 (N=139)	61.5 (N=158)	X^2=4.1, df=1, 0.047
Self-reported race (% white)	87.9 (N=399)	88.3 (N=174)	87.5 (N=225)	X^2=0.1, df=1, 0.885
Education (years)	14.4 (2.8)	14.5 (2.8)	14.3 (2.8)	U=25 758.0, 0.75
CIRS-G burden of physical illness total score	9.8 (4.5)	9.7 (4.6)	9.8 (4.4)	U=24 554.5, 0.582
Diagnosis of hypertension, heart disease or diabetes (% yes)	8.1 (N=37)	9.1 (N=18)	7.4 (N=19)	X^2=0.5, df=1, 0.604
MADRS total score	26.7 (5.8)	26.1 (5.6)	27.1 (5.9)	U=22 802.5, 0.069
MADRS-suicide item score 3 or higher (% yes)	4.8 (N=22)	6.1 (N=12)	3.9 (N=10)	X^2=1.2, df=1, 0.378
BSI-anxiety score	1.5 (0.9)	1.4 (0.9)	1.5 (0.9)	U=23 823.0, 0.281
ATHF (% score ≥3)	61.5 (N=279)	50.3 (N=99)	70.0 (N=180)	X^2=18.4, df=1, <0.001
Duration of current episode (weeks)	293.0 (615.7)	239.9 (557.4)	333.6 (655.1)	U=20 194.0, <0.001
Single or recurrent depression (% recurrent)	71.6 (N=325)	73.6 (N=145)	70.0 (N=180)	X^2=0.7, df=1, 0.463
Age of onset	41.9 (21.5)	42.5 (21.1)	41.4 (21.9)	U=26 296.0, 0.479

All results are presented as N (%) or mean (SD).
*Full response is defined as a decrease of more than 50% in the MADRS baseline score.
†Characteristics of two groups (full responders vs partial and non-responders) are compared using the Mann-Whitney U test for continuous variables and χ^2 test for categorical variables.
ATHF, Antidepressant Treatment History Form; BSI, Brief Symptom Inventory; CIRS-G, Cumulative Illness Rating Scale-Geriatric; MADRS, Montgomery-Asberg Depression Rating Scale.

non-responders (table 2). The comparisons of the proportions of partial responders and non-responders at weeks 4, 6, 8 or 10 who attained full response with additional lengths of treatment showed significant group effects (non-responder vs partial responder) for all weeks (week 4: Wald χ^2=47.7, p<0.001; week 6: Wald χ^2=33.7, p<0.001; week 8: Wald χ^2=29.7, p<0.001, week 10: Wald χ^2=17.3, p<0.001), showing that at week 4 and at each later assessment point, partial responders were more likely than non-responders to become full responders. By contrast, time effects (additional lengths of treatment) were not significant after week 4 (week 4: Wald χ^2=49.9, p<0.001; week 6: Wald χ^2=0.8, p=0.66; week 8: Wald χ^2=1.1, p=0.30; no value for week 10 since only one additional assessment point was available), showing that additional length of treatment on its own does not result in attainment of full response after week 4.

Decision trees predicting eventual treatment response at the early decision point

A total of 333 participants who did not attain full response at the early decision point (ie, week 4) were entered into decision trees. Baseline demographic, clinical variables and week 4 response status (ie, partial vs non-response) were used as potential predictors. In the decision tree minimising false positives (sensitivity threshold=0.3), longer episode duration, having non-response at week 4 and having a previous adequate antidepressant trial predicted not being a full responder at week 12 (figure 1A). In fivefold cross validation, all five training trees included non-response at week 4 as a predictor of not being a full responder at week 12. Having failed to respond to a previous adequate antidepressant trial (four trees), a longer episode duration (four trees), a lower BSI anxiety score (two trees) and more years of education (one tree) also appeared in training trees. NPV of the leftmost terminal node was 84.9%. Average test NPV of the leftmost terminal node was 77.4%. This means that the subgroup of patients who have episode duration of 17 weeks or longer, did not reach partial response at week 4, and had a previous adequate antidepressant trial, have a 77.4% chance of not attaining treatment response after 12 weeks of treatment. Overall, the decision tree had an NPV of 84.9%, PPV of 41.1% and overall accuracy of 56.8%. In fivefold cross-validation, average test NPV of the overall tree was 73.9%, PPV was 35.2% and accuracy was 51.4%. In the decision tree created minimising false negatives (sensitivity threshold=0.7), having non-response at week 4, earlier age at onset, having a previous adequate antidepressant trial, lower baseline MADRS score and longer episode duration predicted not being a full responder at week 12 (figure 1B). In fivefold cross-validation, non-response at week 4 (four trees), having failed to respond to a previous adequate antidepressant trial (three trees), a longer duration of episode (three trees), a lower BSI anxiety score (one tree), being self-reported male (one tree), a lower CIRS-G burden of physical illness total score of less (one tree) and age at onset younger than 25 (one tree) appeared as predictors of not attaining treatment response after 12 weeks of treatment. NPV of the leftmost terminal node was 86.9%. Average test NPV of the leftmost terminal node was 76.6%. This means that the subgroup of patients who did not reach partial treatment response by week 4, had a previous adequate antidepressant trial, and have episode duration of 28 weeks or longer have a 76.6% chance of not attaining treatment response after 12 weeks of treatment. Overall, the decision tree had an NPV of 78.8%, PPV of 45.9% and overall accuracy of 64.8%. In fivefold cross-validation, average test NPV of the overall tree was 69.2%, PPV was 33.5%, and accuracy was 57.1%. Sample size calculation using a method proposed by Riley *et al*[27] for logistic regression was performed post hoc. Minimum sample size that reduces the potential of the developed model to

Table 2 Proportion of full responders, partial responders and non-responders* at each assessment point who attain full response at week 12 (%) and number of full responders, partial responders and non-responders at each assessment point (in brackets)

	Assessment point (week)						
	1	2	4	6	8	10	12
Full responders	**64.9% (37)**	**67.1% (73)**	**75.2% (121)**	**79.0% (143)**	**72.5% (200)**	**73.9%** (207)	**100% (197)**
Partial responders	**57.4% (108)**	**45.2% (135)**	**43.5% (131)**	**43.3% (120)**	36.8% (106)	29.3% (116)	0% (167)
Non-responders	35.9% (309)	35.4% (246)	24.3% (202)	16.8% (191)	8.8% (148)	7.6% (131)	0% (90)

Proportions above 40% are bolded to highlight the groups and time points considered in the selection of the early decision point (see text).
*Full response is defined as a decrease in Montgomery-Asberg Depression Rating Scale (MADRS) score higher than 50% from baseline; partial response as a decrease in MADRS score of 25%–50%; non-response as a decrease in MADRS score of less than 25%

Table 3 (A, B) Percentages of partial responders and non-responders* at each assessment point who attain full response after additional lengths of treatment

A—Partial responders	Assessment points in weeks (N of partial responders)						
Additional lengths of treatment in weeks	1 (N=108)	2 (N=135)	4 (N=131)	6 (N=120)	8 (N=106)	10 (N=116)	Weighted row mean
1	25.9%	–	–	–	–	–	25.9%
2	–	34.8%	32.8%	39.2%	28.3%	29.3%	33.0%
3	48.2%	–	–	–	–	–	48.2%
4	–	37.0%	48.9%	38.3%	36.8%	–	40.5%
5	47.2%	–	–	–	–	–	47.2%
6	–	50.4%	48.9%	43.3%	–	–	47.6%
7	61.1%	–	–	–	–	–	61.1%
8	–	51.9%	43.5%	–	–	–	48.0%
9	60.2%	–	–	–	–	–	60.2%
10	–	45.2%	–	–	–	–	45.2%
11	57.4%	–	–	–	–	–	57.4%
B—Non-responders	**Assessment points in weeks (N of non-responders)**						
Additional lengths of treatment in weeks	1 (N=309)	2 (N=246)	4 (N=202)	6 (N=191)	8 (N=148)	10 (N=131)	Weighted row mean
1	5.2%	–	–	–	–	–	5.2%
2	–	7.7%	5.9%	15.7%	10.1%	7.6%	9.4%
3	13.3%	–	–	–	–	–	13.3%
4	–	14.2%	17.3%	19.4%	8.8%	–	15.1%
5	21.4%	–	–		–	–	21.4%
6	–	28.1%	21.3%	16.8%	–	–	22.5%
7	33.7%	–	–	–	–	–	33.7%
8	–	30.1%	24.3%	–	–	–	27.3%
9	36.3%	–	–	–	–	–	36.3%
10	–	35.4%	–	–	–	–	35.4%
11	35.9%	–	–	–	–	–	35.9%

Weighted row means were calculated using N at each assessment point as weights.
*Full response is defined as a decrease in MADRS score higher than 50% from baseline; partial response as a decrease in MADRS score of 25%–50%; non-response as a decrease in MADRS score of less than 25%.
MADRS, Montgomery-Asberg Depression Rating Scale.

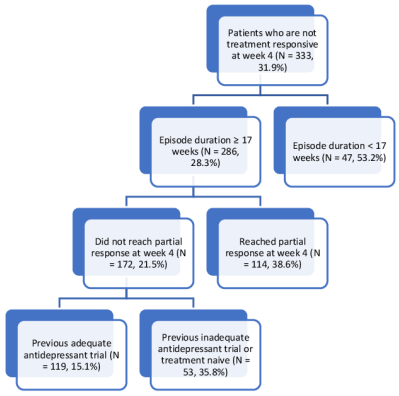
be overfitted to the dataset was found to be 290, which our sample size of 333 exceeds. However, sample size ensuring that the overall prevalence is estimated precisely was calculated to be 384, which exceeds our sample size 333. This corresponds to 13 events per predictor.

DISCUSSION

We analysed data from a 12-week open treatment study of venlafaxine XR in 454 older patients with MDD. We aimed to create decision trees for the early identification of older patients with MDD who are unlikely to respond to their antidepressant. To do this, we first identified the earliest time point where such a decision could be made. Then, we created two decision trees (one to minimise false positives and the other to minimise false negatives) based on baseline demographic and clinical factors to predict the eventual outcome for patients who are partial or non-responders by this early decision point. We selected week 4 as the optimal early decision point; both decision trees for the prediction of participants who would not attain full response after eight additional weeks of treatment shared the following predictors: longer duration of current episode; not having responded to a previous adequate antidepressant trial prior to starting treatment and being a non-responder (vs a partial responder) at week 4. In addition, earlier age at onset and lower depression severity at baseline predicted non-response in the decision tree minimising false negatives (ie, conservatively avoiding falsely predicting non-response). While our findings need to be validated in other larger samples, they show the potential of using baseline clinical characteristics and early symptom improvement to guide clinical decision making. In the future, the use of MBC and decision trees could improve clinical outcomes in patients with difficult-to-treat depression.

In our analysis, less than a quarter of participants who were non-responders after 4 weeks of treatment with venlafaxine XR attained full response at week 12, in contrast to close to half of partial responders. This suggests that early non-responders should be considered for a switch or augmentation as early as week 4, rather than persisting with an antidepressant that is likely to be ineffective. This study both replicates and extends the methods and results of two of our previous studies in older adults with MDD treated with nortriptyline or paroxetine.[4 10] In younger patients with MDD, some[29] but not all studies[22] have suggested that treatment non-response as early as week 2 can be used to predict treatment non-response. However, in our study of older patients with MDD, more than ne-third of non-responders at weeks 1 and 2 attained full response at the end of treatment, suggesting that making a treatment decision before week 4 may result in premature discontinuation of a potentially helpful treatment in too many patients. The difference between our results and some studies in younger patients may also be because some of these studies predicted ultimate response to 6–8 weeks of treatment rather than

A



B

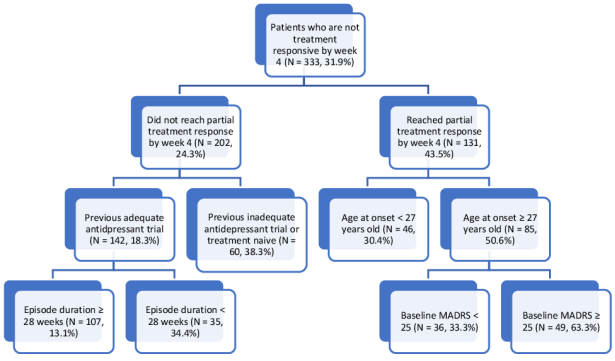


Figure 1 Decision trees predicting full treatment response at week 12 minimising false positives (A) and minimising false negatives (B) for partial and non-responders at week 4. Number of participants belonging to each decision point (eg, n=47 in 'episode duration <17' for figure 1A) and proportion of those participants who attain full response at week 12 are shown in brackets (eg, 53.2% of 47 participants attain full response at week 12 in 'episode duration <17' for figure 1A). Left branches contain predictors associated with lower probability of attaining full response at week 12. MADRS, Montgomery-Asberg Depression Rating Scale.

after a full 12-week course.[22] Both decision trees identified more specific subgroups of participants who are very unlikely to attain treatment response after 12 weeks and should be considered for a change in treatment after 4 weeks. While both trees had NPVs of the leftmost terminal node above 70%, their overall accuracy and PPV were low. The aim of our analysis was to identify the subgroup of patients who would not likely attain eventual treatment response. Thus, clinically, these decision trees would be most useful for identifying participants who are very unlikely to attain eventual treatment response and therefore should be switched to a different treatment.

In our analysis, lower severity of depression (ie, lower baseline MADRS score) was predictive of lower likelihood of attaining full response. This appears to contradict previous findings that a lower severity of depression predicts a higher likelihood of attaining remission.[6 11 19] However, it makes sense that it is harder to reduce a low score by 50% but it is easier to decrease this low score to below a predetermined threshold, as has been discussed in meta-analyses of treatment trials of MDD in both younger[30] and older patients.[6] Lower severity of depression did not appear as a predictor in the training trees in cross-validation, suggesting that it is a weak predictor of not attaining full response.

Our results also replicate several previous studies showing that earlier age of onset of MDD,[31] a longer duration of current episode[6 8 11] and having not responded to at least one adequate prior antidepressant trial[2 11 18] predict non-response. Combining these predictors in a hierarchy by placing them in decision trees operationalise their use for older patients with MDD who did not attain full response by week 4.

Important limitations of this study are the lack of a validating sample and a relatively small sample size equivalent to 9 EPV. While this is close to the rule of thumb of 10 EPV,[26] sample size calculation using Riley and colleagues' method[27] suggested that the minimum sample size for precise estimation of outcome is 384, which is larger than our sample size of 333. Therefore, future studies need to validate our findings in a larger sample. Other limitations of this analysis also include using data from a clinical trial, in which outcomes reflect a systematic treatment process compared with usual care.[32] All participants were treated with venlafaxine XR, and our findings may not generalise to other antidepressants. However, as discussed above, similar findings have been reported with nortriptyline and paroxetine.[4 10] Also, we did not have a placebo group, and the observed symptomatic improvement may have been due to non-specific factors rather than venlafaxine XR. We examined only a limited list of predictors that were shown in previous studies to be predictive of treatment response or remission; this ensured we had sufficient power[33] without performing variable selection, which can complicate data interpretation.[34] Lastly, using a decision tree within a clinical setting may have potential drawbacks, including that it does not account for patient preference. Examining the feasibility and acceptability of using decision trees to guide care would be an important step before they are used in clinical practice.

CLINICAL IMPLICATIONS

Notwithstanding these limitations, methods used in this study can inform future studies using MBC and decision trees to improve the care of older patients with MDD. Specifically, older patients can receive a standard psychiatric assessment including a validated depression scale during their first appointment. Four weeks after initiation of an antidepressant, patients can complete the scale again to determine whether they are full responders, partial responders or non-responders. If they have not attained full response, a shared decision can be made with the patient whether to continue their current antidepressant based on the likelihood it will be helpful predicated on their week 4 response status, age at onset of MDD, episode duration, history of prior antidepressant treatment and baseline symptom severity. Future studies with larger samples need to prospectively compare the outcomes of patients treated following this approach vs those receiving usual care; they may also test the model within specific subgroups, such as among females. Using several different antidepressants would also assess the generalisability of our findings. Furthermore, future studies may also evaluate whether adding biomarkers or digital markers of clinical change[35] into the predictive model may improve our ability to rapidly discern eventual responders versus non-responders, helping to push the field towards true precision medicine in real-world depression care .

Contributors BHM contributed to conducting the study, planning, acquisition of the data, data analysis and reporting of the study. BHM is guarantor. HKK contributed to planning, data analysis and reporting of the study. DMB, JFK, EL and CFR contributed to conduct the study, data acquisition and reporting.

Funding The IRL-GRey study was supported primarily by the National Institute of Mental Health (R01 MH083660 and P30 MH90333 to University of Pittsburgh, R01 MH083648 to Washington University, and R01 MH083643 to University of Toronto). Additional funding was provided by the UPMC Endowment in Geriatric Psychiatry, the Taylor Family Institute for Innovative Psychiatric Research (at Washington University), the Washington University Institute of Clinical and Translational Sciences

grant UL1 TR000448 from the National Centre for Advancing Translational Sciences (NCATS), and the Campbell Family Mental Health Research Institute at the Centre for Addiction and Mental Health, Toronto. Pfizer contributed venlafaxine extended-release capsules for this study and Bristol-Myers Squibb contributed aripiprazole and matching placebo tablets for the randomised phase of the study.

Competing interests HKK has no competing interests to disclose. DMB has received research support from Canadian Institutes of Health Research (CIHR), US National Institute of Health (NIH), Brain Canada, and the Temerty Family through the CAMH Foundation and the Campbell Family Research Institute. He received research support and in-kind equipment support for an investigator-initiated study from Brainsway. He is the site principal investigator for three sponsor-initiated studies for Brainsway. He also receives in-kind equipment support from Magventure for two investigator-initiated studies. He received medication supplies for an investigator-initiated trial from Indivior. JFK received medication supplies from Pfizer and Indivor for investigator-initiated studies. He is also on the scientific advisory board of Aifred Health. EL has received funding from Takeda, Lundbeck, Janssen, Alkermes, Aptynx, and Patient-Centered Outcomes Research Institute (PCORI). He is also a consultant for Janssen and Jazz Pharmaceuticals. CFR received research support from NIH, PCORI, American foundation for suicide prevention, center for Medicare and Medicaid services, commonwealth of Pennsylvania, and the Brain and behavior research foundation. His work is also sponsored by pharmaceutical supplies from Bristol Meyers Squib and Pfizer. BHM receives support from the Labatt Family Chair in Biology of Depression in Late-Life Adults at the University of Toronto. He receives research support from Brain Canada, CIHR, the CAMH Foundation, PCORI, NIH, Capital Solution Design LLC (software used in a study founded by CAMH Foundation), and HAPPYneuron (software used in a study founded by Brain Canada). He has been an unpaid consultant to Myriad Neuroscience.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request.

ORCID iD

Helena Kyunghee Kim http://orcid.org/0000-0002-6578-2109

REFERENCES

1 Zhu M, Hong RH, Yang T, *et al*. The efficacy of Measurement-Based care for depressive disorders: systematic review and meta-analysis of randomized controlled trials. *J Clin Psychiatry* 2021;82. doi:10.4088/JCP.21r14034. [Epub ahead of print: 28 09 2021].
2 Buchalter ELF, Oughli HA, Lenze EJ, *et al*. Predicting remission in late-life major depression: a clinical algorithm based upon past treatment history. *J Clin Psychiatry* 2019;80. doi:10.4088/JCP.18m12483. [Epub ahead of print: 10 12 2019].
3 Coley RY, Boggs JM, Beck A, *et al*. Defining success in Measurement-Based care for depression: a comparison of common metrics. *Psychiatr Serv* 2020;71:312–8.
4 Mulsant BH, Houck PR, Gildengers AG, *et al*. What is the optimal duration of a short-term antidepressant trial when treating geriatric depression? *J Clin Psychopharmacol* 2006;26:113–20.
5 Dew MA, Whyte EM, Lenze EJ, *et al*. Recovery from major depression in older adults receiving augmentation of antidepressant pharmacotherapy. *Am J Psychiatry* 2007;164:892–9.
6 Tunvirachaisakul C, Gould RL, Coulson MC, *et al*. Predictors of treatment outcome in depression in later life: a systematic review and meta-analysis. *J Affect Disord* 2018;227:164–82.
7 Andreescu C, Lenze EJ, Dew MA, *et al*. Effect of comorbid anxiety on treatment response and relapse risk in late-life depression: controlled study. *Br J Psychiatry* 2007;190:344–9.
8 Kautzky A, Dold M, Bartova L, *et al*. Clinical factors predicting treatment resistant depression: affirmative results from the European multicenter study. *Acta Psychiatr Scand* 2019;139:78–88.
9 Driscoll HC, Basinski J, Mulsant BH, *et al*. Late-Onset major depression: clinical and treatment-response variability. *Int J Geriatr Psychiatry* 2005;20:661–7.
10 Andreescu C, Mulsant BH, Houck PR, *et al*. Empirically derived decision trees for the treatment of late-life depression. *Am J Psychiatry* 2008;165:855–62.
11 Joel I, Begley AE, Mulsant BH, *et al*. Dynamic prediction of treatment response in late-life depression. *Am J Geriatr Psychiatry* 2014;22:167–76.
12 Lenze EJ, Mulsant BH, Blumberger DM, *et al*. Efficacy, safety, and tolerability of augmentation pharmacotherapy with aripiprazole for treatment-resistant depression in late life: a randomised, double-blind, placebo-controlled trial. *Lancet* 2015;386:2404–12.
13 American Psychiatric Association. *Diagnostic and statistical manual of mental disorders : DSM-IV*. Washington, DC: American Psychiatric Association, 1994.
14 Ventura J, Liberman RP, Green MF, *et al*. Training and quality assurance with the structured clinical interview for DSM-IV (SCID-I/P). *Psychiatry Res* 1998;79:163–73.
15 Davidson J, Turnbull CD, Strickland R, *et al*. The Montgomery-Asberg depression scale: reliability and validity. *Acta Psychiatr Scand* 1986;73:544–8.
16 Miller MD, Paradis CF, Houck PR, *et al*. Rating chronic medical illness burden in geropsychiatric practice and research: application of the cumulative illness rating scale. *Psychiatry Res* 1992;41:237–48.
17 Derogatis LR, Melisaratos N. The brief symptom inventory: an introductory report. *Psychol Med* 1983;13:595–605.
18 Hsu JH, Mulsant BH, Lenze EJ, *et al*. Clinical predictors of extrapyramidal symptoms associated with aripiprazole augmentation for the treatment of late-life depression in a randomized controlled trial. *J Clin Psychiatry* 2018;79. doi:10.4088/JCP.17m11764. [Epub ahead of print: 19 06 2018].
19 Gildengers AG, Houck PR, Mulsant BH, *et al*. Trajectories of treatment response in late-life depression: psychosocial and clinical correlates. *J Clin Psychopharmacol* 2005;25:S8–13.
20 Jackson WC, Papakostas GI, Rafeyan R, *et al*. Recognizing inadequate response in patients with major depressive disorder. *J Clin Psychiatry* 2020;81. doi:10.4088/JCP.OT19037BR2. [Epub ahead of print: 05 05 2020].
21 Tedeschini E, Levkovitz Y, Iovieno N, *et al*. Efficacy of antidepressants for late-life depression: a meta-analysis and meta-regression of placebo-controlled randomized trials. *J Clin Psychiatry* 2011;72:1660–8.
22 de Vries YA, Roest AM, Bos EH, *et al*. Predicting antidepressant response by monitoring early improvement of individual symptoms of depression: individual patient data meta-analysis. *Br J Psychiatry* 2019;214:4–10.
23 Olgiati P, Serretti A, Souery D, *et al*. Early improvement and response to antidepressant medications in adults with major depressive disorder. meta-analysis and study of a sample with treatment-resistant depression. *J Affect Disord* 2018;227:777–86.
24 Hubbard AE, Ahern J, Fleischer NL, *et al*. To Gee or not to Gee: comparing population average and mixed models for estimating the associations between neighborhood risk factors and health. *Epidemiology* 2010;21:467–74.
25 Collins GS, Reitsma JB, Altman DG, *et al*. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ* 2015;350:g7594.
26 Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and COX regression. *Am J Epidemiol* 2007;165:710–8.
27 Riley RD, Snell KI, Ensor J, *et al*. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. *Stat Med* 2019;38:1276–96.
28 Kiernan M, Kraemer HC, Winkleby MA, *et al*. Do logistic regression and signal detection identify different subgroups at risk? implications for the design of tailored interventions. *Psychol Methods* 2001;6:35–48.
29 Wagner S, Engel A, Engelmann J, *et al*. Early improvement as a resilience signal predicting later remission to antidepressant treatment in patients with major depressive disorder: systematic review and meta-analysis. *J Psychiatr Res* 2017;94:96–106.
30 Bower P, Kontopantelis E, Sutton A, *et al*. Influence of initial severity of depression on effectiveness of low intensity interventions: meta-analysis of individual patient data. *BMJ* 2013;346:f540.
31 Kautzky A, Baldinger-Melich P, Kranz GS, *et al*. A new prediction model for evaluating treatment-resistant depression. *J Clin Psychiatry* 2017;78:215–22.
32 Mulsant BH, Blumberger DM, Ismail Z, *et al*. A systematic approach to pharmacotherapy for geriatric major depression. *Clin Geriatr Med* 2014;30:517–34.
33 Ogundimu EO, Altman DG, Collins GS. Adequate sample size for developing prediction models is not simply related to events per variable. *J Clin Epidemiol* 2016;76:175–82.
34 Heinze G, Dunkler D. Five myths about variable selection. *Transpl Int* 2017;30:6–10.
35 Mofsen AM, Rodebaugh TL, Nicol GE, *et al*. When all else fails, listen to the patient: a viewpoint on the use of ecological Momentary assessment in clinical trials. *JMIR Ment Health* 2019;6:e11845.