Treating Depression After Initial Treatment Failure

*Directly Comparing Switch and Augmenting Strategies in STAR*D*

Bradley N. Gaynes, MD, MPH, Stacie B. Dusetzina, PhD,† Alan R. Ellis, MSW,‡ Richard A. Hansen, PhD,§|| Joel F. Farley, PhD,|| William C. Miller, MD, PhD,¶ and Til Stürmer, MD, MPH¶*

Objective: Augmenting and switching antidepressant medications are the 2 most common next-step strategies for depressed patients failing initial medication treatment. These approaches have not been directly compared; thus, our objectives are to compare outcomes for medication augmentation versus switching for patients with major depressive disorder in the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) clinical trial.

Methods: We conducted a retrospective analysis of participants aged 18 to 75 years with DSM-IV nonpsychotic depression who failed to remit with initial treatment in the STAR*D clinical trial (N = 1292). We compared depressive symptom remission, response, and quality of life among participants in each study arm using propensity score matching to minimize selection bias.

Results: The propensity-score-matched augment (N = 269) and switch (N = 269) groups were well balanced on measured characteristics. Neither the likelihood of remission (risk ratio, 1.14; 95% confidence level, 0.82–1.58) or response (risk ratio, 1.14; 95% confidence level, 0.82–1.58), nor the time to remission (log-rank test, $P = 0.946$) or response (log-rank test, $P = 0.243$) differed by treatment strategy. Similarly, quality of life did not differ. Post hoc analyses suggested that augmentation improved outcomes for patients tolerating 12 or more weeks of initial treatment and those with partial initial treatment response.

Conclusions: For patients receiving and tolerating aggressive initial antidepressant trials, there is no clear preference for next-step augmentation versus switching. Findings tentatively suggest that those who complete an initial treatment of 12 weeks or more and have a partial response with residual mild depressive severity may benefit more from augmentation relative to switching.

Key Words: major depressive disorder, treatment resistance, propensity score, outcome

From the *Department of Psychiatry, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC; †Department of Health Care Policy, Harvard Medical School, Boston, MA; ‡Cecil G. Sheps Center for Health Services Research, University of North Carolina at Chapel Hill, Chapel Hill, NC; §Harrison School of Pharmacy, Auburn University, Auburn, AL; ||Division of Pharmaceutical Outcomes and Policy, Eshelman School of Pharmacy, and ¶Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC.

Received December 21, 2010; accepted after revision July 26, 2011.

Reprints: Bradley N. Gaynes, MD, MPH, Department of Psychiatry, School of Medicine, University of North Carolina at Chapel Hill, CB# 7160 Chapel Hill, NC 27599 (e-mail: bradley_gaynes@med.unc.edu).

This project was funded under contract no. 290200500401 from the Agency for Healthcare Research and Quality (AHRQ), US Department of Health and Human Services, as part of the Developing Evidence to Inform Decisions about Effectiveness program. The authors of this report are responsible for its content. Although the funders provided feedback on the contract application and made decisions about whether to fund (as would happen with an application for a National Institutes for Health [NIH]-funded grant), the funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Statements in the report should not be construed as endorsement by the AHRQ or the US Department of Health and Human Services.

ISSN: 0271-0749
DOI: 10.1097/JCP.0b013e31823f705d

(*J Clin Psychopharmacol* 2012;32: 114–119)

Major depressive disorder (MDD) is projected to be the second leading cause of disability worldwide by 2020.[1] Although existing antidepressant treatments are often effective, as many as 40% of depressed patients fail to respond after first-line treatment.[2] Of greater concern, only one-third of patients fully recover, or remit, with the initial treatment attempt.[3] Improving clinical remission rates early in the course of treatment could significantly improve the quality of care for patients with depression.

However, limited evidence is available to guide clinical decision making after initial treatment failure. Adding an additional medication to augment initial treatment and switching to a different antidepressant are the 2 most common secondary treatment strategies, but these approaches have not been directly compared. The Sequenced Treatment Alternatives to Relieve Depression (STAR*D) trial, a large-scale effectiveness clinical trial, was designed to provide evidence on preferred treatment of MDD after initial treatment failure.[4,5] Among patients who did not achieve remission with first-line citalopram treatment (level 1), approximately one-third of patients augmenting treatment and one-quarter of patients switching treatment achieved subsequent remission (level 2).

To mimic real-world care, STAR*D used an equipoise randomization scheme for treatment assignment,[6] where patient choice could influence treatment selection. In STAR*D, patients had clear preferences for treatment strategies, and patients who received augmentation differed from patients who switched antidepressants.[7,8] To date, the augmentation and switch arms of the trial have not been directly compared, leaving clinicians uncertain about next steps for patients who fail initial antidepressant treatment.[7] A fundamental question remains: when a patient has failed a trial of medication at an adequate dose and for a sufficient period, should the physician augment the first medication or switch to a new drug?

To address this issue, we used a nonexperimental design to reanalyze the existing STAR*D trial data, applying propensity scoring to minimize imbalances in patient, disease, and treatment characteristics that precluded an augmentation-versus-switch comparison in the original study. We used the public access STAR*D data set to address 3 questions:

1. Does the likelihood of or time to remission differ as a function of treatment strategy?
2. Does the likelihood of or time to response differ as a function of treatment strategy?
3. Does quality of life differ as a function of treatment strategy?

Among patients who tolerate an initial medication trial but only partially respond, clinical practice suggests that augmentation is favored over switching treatments.[9] Therefore, we hypothesized that augmentation would remain the preferred strategy compared with switching after adjustment.

MATERIALS AND METHODS

Study Overview

The original STAR*D trial methods, design, and rationale have been detailed previously.[4,5] Briefly, STAR*D was designed to assess which treatments would be most effective for outpatients with nonpsychotic MDD who had unsatisfactory responses to initial treatments. Its inclusive selection criteria enrolled participants with multiple comorbid medical and psychiatric illnesses. All participants were established clinic patients, and many (40%) were from primary care settings. The distribution of depressive severity was similar to nationally representative samples, and the racial and ethnic composition of enrolled patients approximated the US census.

Eligible participants who consented to STAR*D were initially given citalopram. Participants were required to continue citalopram for 12 weeks unless (1) intolerable adverse effects required medication change, (2) optimal dose increase was not possible because of adverse effects or participants' choice, or (3) significant symptoms were present after 9 weeks at maximally tolerated dose. Participants meeting these criteria were eligible to enter a series of non–placebo-controlled randomized clinical trials. This analysis will address the next-step treatments, known as level 2 treatments because they followed failure of the first step.

Study Population

There were 4041 participants aged 18 to 75 years with nonpsychotic depression enrolled in STAR*D. Of 4041 entering initial treatment, 1439 participants did not achieve remission and entered level 2. Excluding those who received cognitive behavioral therapy as a second-step treatment (n = 147), we focused on participants who received treatment within the level 2 medication arms (N = 1292). The efficacy of alternative augment strategies tested (citalopram + bupropion SR, n = 279; citalopram + buspirone, n = 286) did not differ, nor did those of alternative switch strategies (bupropion, n = 239; sertraline, n = 238; and venlafaxine XR, n = 250).[7] Accordingly, we combined augment and switch strategies, respectively, into a switch (N = 727) and an augment (N = 565) group.

Measures

Within each treatment level, visits were conducted at baseline and weeks 2, 4, 6, 9, and 12, with an additional visit at week 14 if needed. At baseline, research coordinators collected demographic information, self-reported psychiatric history, and current general medical conditions (using the Cumulative Illness Rating Scale[10,11]). Additionally, they administered the Hamilton Rating Scale for Depression ($HRSD_{17}$),[12,13] and patients completed a self-report version of the 16-item Quick Inventory of Depressive Symptomatology, Self-Rated ($QIDS\text{-}SR_{16}$).[14,15] At each follow-up visit, research coordinators collected $QIDS\text{-}SR_{16}$ and adverse effects ratings (using the Frequency, Intensity, and Burden of Side Effects Rating [FIBSER], with higher ratings indicating greater severity).[5] At the exit from each treatment level, measures including the $HRSD_{17}$ were collected by outcome assessors, masked to treatment group. Interactive Voice Response assessments of clinical symptoms, adverse effects, and self-reported quality of life were conducted quarterly.

Outcomes

Remission was defined as a score of 5 or lower on the $QIDS\text{-}SR_{16}$ at level 2 exit (scores range from 0 to 27, higher scores represent greater symptom severity). The secondary remission outcome was defined as a score of 7 or lower on the $HRSD_{17}$ (scores range from 0 to 52, with higher scores representing greater symptom severity). Per the original STAR*D protocol,[5] participants for whom the level 2 exit $HRSD_{17}$ score was missing were designated as not achieving remission. The $QIDS\text{-}SR_{16}$ and $HRSD_{17}$ scores correlate closely[15] and can be used to produce equivalent scores. However, the $QIDS\text{-}SR_{16}$ was selected as the primary remission outcome as it was more likely than the $HRSD_{17}$ to be collected at a time close to a patient's exit from a level.

Response was defined as a decrease of 50% or more in $QIDS\text{-}SR_{16}$ score from level 2 entry to level 2 exit.

Quality of life was measured using the physical health and mental health subscales of the Short-Form Health Survey (SF-12) and the Quality of Life Enjoyment and Satisfaction Questionnaire (Q-LES-Q). The 2 SF-12 subscales range from 0 to 100, with higher scores indicating better functioning. The population norm for each is 50 ± 10.[16] The 16-item Q-LES-Q measures the degree of enjoyment and satisfaction experienced by subjects in daily functioning.[17,18] The total score is based on responses to the first 14 items on the scale (each scored on a 5-point Likert scale with 1 = very poor, 5 = very good). Similar to the SF-12, higher scores represent greater life enjoyment and satisfaction.

Statistical Analysis

We examined missing data patterns for variables in the propensity score and outcome models. The level of missing information in the analysis data set was only 5%, but 82% of observations had at least 1 missing value. To replace the missing values, we generated 20 imputed data sets using the Markov chain Monte Carlo method in SAS PROC MI.[19,20]

Propensity scores were developed using multiple logistic regression to estimate each participant's conditional probability of receiving medication augmentation. Forty-seven covariates were selected for inclusion in the models based on their relationship with the treatment and the outcome (or the outcome alone). Thirty interaction terms were selected using a forward stepwise selection process that minimized covariate imbalance. We individually matched patients from the switch group to patients in the augment group based on their estimated propensity scores, resulting in the selection of 269 augment participants, matched with an equal number of switch participants.

On the basis of the propensity-score-matched samples, we used binomial regression models to measure the association between medication treatment arm (augment vs switch) and remission and response rates. We generated incidence proportions (IPs) and risk ratios (RRs) for remission and response using PROC GENMOD.[21]

Time to first remission and response were analyzed using survival analysis. Kaplan-Meier curves were generated and a log-rank statistic was used to compare the curves. Cox proportional hazards models were used to estimate hazard ratios. For the survival analysis, we used a singly imputed data set based on the expectation-maximization method.[19,20]

We used linear regression to estimate the differences between the switch and augment groups in mean SF-12 physical health score, SF-12 mental health score, and Q-LES-Q total score.

RESULTS

Before propensity scoring, the augmentation group differed from the switch group in several ways (Table 1). Augmenters had fewer lifetime depressive episodes (5.6 vs 6.8), shorter current depressive episode duration (27.8 vs 30.3 months), a greater proportion with drug (8.7% vs 5.5%) or alcohol abuse (13.0% vs 11.9%), and fewer with posttraumatic stress disorder (18.9% vs 23.0%) compared to switchers. Upon exiting level 1 treatment, augmenters had lower depressive severity by $QIDS\text{-}SR_{16}$ (11.4 vs 13.2)

TABLE 1. Participant Characteristics Before and After Propensity Score Matching

	Before Propensity Score Matching		After Propensity Score Matching		Unmatched Patients	
	Augment (N = 565)	Switch (N = 727)	Augment (N = 269)	Switch (N = 269)	Augment (N = 296)	Switch (N = 458)
Demographic characteristics						
Age, y	41.6 (12.7)	42.4 (12.8)	41.7 (12.8)	41.8 (12.9)	41.5 (12.6)	42.7 (12.7)
Female sex, %	58.8	58.7	59.0	58.4	58.5	58.9
Race, %						
White	78.0	75.8	77.4	77.0	78.6	75.1
Black	18.0	18.7	18.4	18.9	17.7	18.6
Education, %						
More than high school	22.3	23.1	20.4	21.2	24.1	24.2
Employment status, %						
Employed	64.0	63.1	65.1	64.7	63.0	62.2
Marital status, %						
Married or cohabitating	43.8	43.0	44.7	44.4	43.0	42.2
Divorced, separated or widowed	28.2	30.7	28.6	28.7	27.8	31.8
Psychiatric history (baseline)						
No. major depressive episodes	5.6 (8.8)	6.8 (11.1)	6.5 (10.4)	6.3 (10.1)	4.8 (7.0)	7.1 (11.6)
Duration of index depressive episode						
Mean, mo	27.8 (56.1)	30.3 (66.7)	25.8 (52.0)	25.9 (58.1)	29.6 (59.6)	33.0 (71.1)
≥2 y, %	27.6	27.6	26.2	25.7	28.9	28.7
Psychiatric comorbidities, %						
Drug abuse	8.7	5.5	7.0	7.1	10.4	4.6
Alcohol abuse	13.0	11.9	11.7	11.9	14.2	12
Posttraumatic stress disorder	18.9	23.0	21.5	20.6	16.5	24.4
Generalized anxiety disorder	23.4	22.2	23.0	22.8	23.8	21.9
Level 1 exit characteristics						
Depressive severity:						
$QIDS\text{-}SR_{16}$ score	11.4 (4.9)	13.2 (4.9)	12.5 (5.1)	12.6 (4.9)	10.3 (4.5)	13.6 (4.9)
$HRSD_{17}$ score	16.0 (7.1)	19.0 (7.3)	17.6 (7.2)	17.6 (7.3)	14.6 (6.7)	19.9 (7.2)
SF-12						
Mental	32.4 (9.2)	29.7 (9.6)	31.0 (8.9)	30.8 (9.4)	33.7 (9.2)	29.1 (9.6)
Physical	46.9 (11.8)	45.6 (12.0)	46.1 (12.1)	46.2 (11.8)	47.7 (11.4)	45.3 (12.2)
Q-LES-Q score	45.7 (16.3)	41.1 (16.9)	43.0 (16.5)	43.1 (16.5)	48.1 (15.7)	39.9 (17.0)
Adverse effect measures						
FIBSER score	1.2 (1.3)	2.8 (1.9)	1.4 (1.4)	1.4 (1.5)	1.0 (1.2)	3.7 (1.6)
Exited level 1 because of adverse effects, %	10.3	62.8	19.9	20.2	1.6	87.8
Experienced a serious adverse event, %	4.6	2.6	3.5	3.5	5.6	2.1
Management features						
Duration of level 1 treatment, wk	11.9 (2.9)	8.0 (4.2)	10.9 (3.2)	10.9 (3.0)	12.8 (2.4)	6.3 (3.8)
Citalopram dose, mg/d	55.1 (10.9)	41.4 (17.7)	52.8 (12.8)	52.9 (12.4)	57.1 (8.4)	34.7 (16.8)
Medication adherence, %						
Never/rarely missing	89.6	82.6	88.9	87.4	90.1	79.7

Data are means (SD).

Level 1 refers to initial treatment with citalopram. Level 2 refers to second-step treatment with either medication augmentation or switching.

$QIDS\text{-}SR_{16}$ indicates the 16-item Quick Inventory of Depressive Symptomatology, Self-Rated (scores can range from 0 to 27; higher scores indicate increased severity of depressive symptoms); $HRSD_{17}$, the 17-item Hamilton Rating Scale for Depression (scores can range from 0 to 52; higher scores indicate increased severity of depressive symptoms); SF-12, Short-Form Health Survey, 2 SF-12 subscales (mental and physical) range from 0 to 100; higher scores indicate increased perceived functioning. The population norm for each is 50 ± 10. Q-LES-Q, Quality of Life Enjoyment and Satisfaction Questionnaire (scores can range from 0 to 100; higher scores indicate greater satisfaction); FIBSER is a global measure of the intensity and the burden of adverse effects. Scores were averaged to calculate a composite score.

and slightly higher quality of life by SF-12 and Q-LES-Q measures. Furthermore, augmenters were less likely to exit level 1 because of adverse effects (10.3% vs 62.8%), spent more time in level 1 treatment (11.9 vs 8.0 weeks), and had higher citalopram daily doses at level 1 exit (55.1 vs 41.4 mg) than switchers.

TABLE 2. Comparison of Remission, Response, and Quality of Life by Treatment Strategy: Full Study Sample and Propensity-Score-Matched Sample

	Unadjusted Analysis—Full Study Sample					
	Augment (N = 565)		Switch (N = 727)		Augment vs Switch (N = 1292)	
	IP	95% CI	IP	95% CI	RR	95% CI
Measures of remission						
QIDS-SR_{16}	0.33	0.29–0.37	0.23	0.20–0.26	1.47	1.23–1.75
$HRSD_{17}$	0.28	0.25–0.32	0.20	0.17–0.23	1.40	1.15–1.70
Measures of response						
QIDS-SR_{16}	0.27	0.23–0.30	0.24	0.20–0.27	1.14	0.94–1.37
Measures of Quality of Life	**Mean**	**95% CI**	**Mean**	**95% CI**	**Mean Difference**	**95% CI**
Q-LES-Q	53.3	50.9–55.6	50.9	48.9–53.0	2.3	−0.8 to 5.5
SF-12						
Mental health status	40.2	38.7–41.7	38.0	36.8–39.3	2.2	0.3–4.1
Physical health status	45.9	44.5–47.3	45.1	43.9–46.3	0.8	−1.1 to 2.7
	Adjusted Analysis—Propensity-Score-Matched Sample					
	Augment (N = 565)		Switch (N = 727)		Augment vs Switch (N = 1292)	
	IP	95% CI	IP	95% CI	RR	95% CI
Measures of remission						
QIDS-SR_{16}	0.28	0.21–0.34	0.24	0.19–0.30	1.14	0.82–1.58
$HRSD_{17}$	0.24	0.18–0.29	0.22	0.16–0.28	1.07	0.76–1.50
Measures of response						
QIDS-SR_{16}	0.27	0.21–0.33	0.23	0.18–0.29	1.14	0.82–1.58
Measures of Quality of Life	**Mean**	**95% CI**	**Mean**	**95% CI**	**Mean Difference**	**95% CI**
Q-LES-Q	49.7	45.8–53.6	50.8	47.0–54.6	−1.1	−6.6 to 4.4
SF-12						
Mental health status	38.0	35.3–40.6	37.6	35.3–39.9	0.4	−3.0 to 3.7
Physical health status	45.1	42.6–47.6	45.5	43.3–47.8	−0.4	−3.8 to 2.9

Remission on the QIDS-SR_{16} is defined as a level 2 exit score of 5 or less. Remission on the $HRSD_{17}$ is defined as a level 2 exit score of 7 or less. Response was defined as a decrease of 50% or more in QIDS-SR_{16} score from level 2 entry to level 2 exit. QIDS-SR_{16} indicates the 16-item Quick Inventory of Depressive Symptomatology, Self-Rated (scores can range from 0 to 27; higher scores indicate increased severity of depressive symptoms); $HRSD_{17}$, the 17-item Hamilton Rating Scale for Depression (scores can range from 0 to 52; higher scores indicate increased severity of depressive symptoms); Q-LES-Q, Quality of Life Enjoyment and Satisfaction Questionnaire (scores can range from 0 to 100; higher scores indicate greater satisfaction); SF-12, Short-Form Health Survey, 2 SF-12 subscales (mental and physical) range from 0 to 100; higher scores indicate increased perceived functioning. The population norm for each is 50 ± 10.

The propensity score matching corrected for these important differences between augmentation and switch groups in level 2, resulting in a smaller but well-balanced sample. The matched sample included an average of 269 patients in each group (augment and switch) across the 20 imputations (Table 1). This sample had a mean age of approximately 42 years and was primarily female (60%), white (~77%), and employed (~65%). Overall, sociodemographic characteristics were similar before and after propensity score matching.

Patients excluded from the matched sample had distinct characteristics that differed as function of whether they received augmentation or switched treatments (Table 1). Those excluded who received augmentation tended to be less severely depressed, remained in level 1 treatment longer, received a higher dose of citalopram during level 1 treatment, and were less likely to exit level 1 because of adverse effects than were excluded patients who received switch treatment.

Despite an unadjusted RR in the full data set favoring augmentation (RR, 1.47; 95% confidence level [CI], 1.23–1.75), after matching the likelihood of remission (QIDS-SR_{16} ≤5) did not differ between groups (RR, 1.14; 95% CI, 0.82–1.58) (Table 2). Similarly, the likelihood of remission using the $HRSD_{17}$ did not differ. Furthermore, the time to remission was similar between augmentation and switch strategies (hazard ratio, 1.01; 95% CI, 0.76–1.34; $P = 0.946$).

Similar to the unadjusted estimates of response by QIDS-SR_{16}, in the matched analysis, treatment strategy did not significantly affect the likelihood of response (RR, 1.14; 95% CI, 0.82–1.58). Additionally, treatment strategy did not affect the time to response (hazard ratio, 0.85; 95% CI, 0.64–1.12; $P = 0.243$).

Finally, treatment strategy did not affect quality of life, as measured by either the Q-LES-Q or the SF-12 Mental Health and Physical Health subscales.

In STAR*D, the mean doses of switch medications were lower than the adequate doses of the medications defined in the study protocol.[22] Therefore, we assessed whether dosing affected our results. We created variables to indicate each patient's dose(s) based on standard dosing ranges[23] and assessed dose effects. In a sensitivity analysis, we adjusted for dosing between

groups; none of the dose variables had a meaningful effect on any of the outcomes.

Post Hoc Analyses

Although outcomes did not differ in the matched sample overall, the treatment effect may have been heterogeneous. Propensity score stratification facilitates the assessment of treatment effect heterogeneity through comparisons across strata. Stratification controls for pretreatment differences by matching several groups on the propensity score rather than creating 1:1 matches between individuals. Using the full sample, we assessed whether the likelihood of remission differed by the propensity to receive augmentation by stratifying into quintiles, from low (most of this group switched) to high (most of this group augmented) propensity to receive augmentation.

Those with the highest propensity for augmentation were 1.65 times as likely to remit with augmentation as with switching (95% CI, 0.86–3.16), suggesting a possible benefit of augmentation over switching for this subgroup of patients. Those with the lowest propensity for augmentation were also more likely to remit with augmentation (RR, 2.42), but that group contained few augmenting participants (n = 4) and the estimate was imprecise (95% CI, 0.73–8.05).

To explore the reasons for this heterogeneity and inform clinicians' treatment decisions, we identified 2 clinically relevant variables that strongly influenced the propensity to receive augmentation: depressive severity and number of weeks in first step treatment. We stratified on each variable to assess its effect on the likelihood of remission with augmentation versus switch. Comparisons made within strata were based on the full study sample and were unadjusted for other covariates.

Time on initial therapy consistently increased the likelihood of remission with augmentation; however, the effect was strongest for patients who received 12 weeks of initial therapy. For these patients, augmentation was nearly twice as likely to produce remission as was switch (12–13 weeks, augment n = 164, switch n = 89; unadjusted RR, 1.9; 95% CI, 1.16–3.11). Depressive severity after initial therapy also demonstrated a consistent trend with increased likelihood of remission with augmentation (as compared with switching) for those with lower depressive severity. This effect was strongest for patients with residual depressive severity scores less than 9 (augment n = 175, switch n = 119; unadjusted RR, 1.32; 95% CI, 1.03–1.70).

DISCUSSION

For patients who received an aggressive initial antidepressant trial and tolerated it well, there is no clear preference for either switching or augmenting as a next-step treatment. Our analyses provide the first direct comparison of the benefit of augmentation versus switch strategies for the large proportion of patients whose depression does not remit with the initial treatment attempt. Using propensity score matching, we were able to minimize imbalances in patient, disease, and treatment characteristics to test directly whether augmentation produced better clinical outcomes than switching after failure of an initial antidepressant treatment of MDD. Although the unadjusted results suggested greater benefit from augmentation, the propensity-score-matched comparison, which accounted for measured differences between the groups, showed no greater clinical benefit from choosing augmentation over switching. Contrary to our predictions of better outcomes for augmentation, we found that the likelihood of remission or response, time to remission or response, and quality of life measures did not differ by treatment strategy.

Importantly, the matched sample does not reflect the full STAR*D trial population, but reflects patients who were moderately depressed, able to tolerate an aggressive (high dose) level 1 treatment for approximately 11 weeks, and tended not to discontinue initial treatment because of adverse effects. Patients who were excluded from the matched sample had distinct characteristics that differed as a function of whether they received augmentation or switch treatments. Specifically, within the unmatched sample, the switch group (compared to the augmentation group) had larger proportions of patients intolerant of citalopram and patients with greater depressive severity. Accordingly, our matched sample results apply to patients who received and were tolerant of a well-dosed acute phase depression trial and who were assigned to either augmenting or switching in sufficient numbers. For such a group, the selection of a switching versus augmenting strategy did not change the likelihood of a better clinical outcome. However, based on post hoc analyses of the full STAR*D trial population, augmentation produced better outcomes for patients tolerating 12 or more weeks of level 1 treatment and who had partially responded to initial treatment (mild depressive severity, QIDS-SR_{16} <9). For patients fitting this profile, augmentation may be preferred to switch.

There were limitations to our analyses. Although propensity score matching minimized the selection bias introduced by the original STAR*D trial design, it did so at the cost of statistical power and generalizability. First, the group overlap was not ideal, leading to a smaller sample size with our matched analysis. As a result, our matched sample represents only a subset of the full STAR*D trial population. Second, it is difficult to make conclusions about differential treatment effects from the propensity-score-stratified analysis because of limited sample size. We also stratified on 2 clinically relevant variables that strongly influenced the propensity score; conclusions from those analyses are not as limited by sample size but are subject to residual confounding within strata. Third, remission and response were uncommon in our propensity-score-matched sample, resulting in limited power to detect a treatment effect. Fourth, our propensity score was restricted to those variables that were measured during the original trial. Although the STAR*D study collected many variables indicative of the likelihood of depression remission, unmeasured confounding still may exist. Finally, the analysis included only a limited number of augmentation and switch options which, at the time the study was designed, represented treatment options with the strongest scientific support. The use of other augmentation strategies, such as atypical antipsychotic medications,[24] different second-step antidepressants,[25,26] or different medication combinations[27] might produce different results.

For many patients not tolerating an initial antidepressant treatment, the choice is clear: stop the offending medication and switch to a different antidepressant. However, the choice is less clear for those who tolerate but do not fully recover after the initial treatment. For these patients, clinicians need evidence-based guidance about next-step selection. Our propensity-matched sample represents that group. Our analyses indicate that for these patients, augmentation versus switch strategies yield comparable outcomes, whether measured by remission, response, or quality of life. However, our results also provide some tentative evidence that for patients who tolerate initial medication trials for 12 weeks and have depressive severity scores indicating a partial response, clinicians may consider recommending augmentation over switch.

AUTHOR DISCLOSURE INFORMATION

The authors have no conflicts of interest directly related to the content of this study. The following financial relationships might be construed as conflicts of interest. Dr Gaynes has received

research or consulting support from AHRQ, NIH, Bristol-Myers Squibb, Novartis, M-3 Information, and Shire Pharmaceuticals. Dr Dusetzina was supported by AHRQ grant 5-T-32 HS000032-20. Dr Stürmer and Mr Ellis have received funding from the UNC-GlaxoSmithKline Center for Excellence in Pharmacoepidemiology and Public Health, which is supported by unrestricted research grants from pharmaceutical companies to UNC. Dr Stürmer also has received research support from AHRQ and NIH. Dr Hansen has received research or consulting support from AHRQ, NIH, Takeda Pharmaceuticals, and Novartis. Dr Farley has received unrestricted grant support from the Pfizer Foundation and Robert Wood Johnson Foundation and consulting fees from Takeda Pharmaceuticals and Novartis. Dr Miller has received research support from AHRQ and NIH.

REFERENCES

1. Murray CJ, Lopez AD. Alternative projections of mortality and disability by cause 1990–2020: Global Burden of Disease Study. *Lancet*. 1997;349(9064):1498–1504.
2. Gartlehner G, Gaynes BN, Hansen RA, et al. Comparative benefits and harms of second-generation antidepressants: background paper for the American College of Physicians. *Ann Intern Med*. 2008;149(10):734–750.
3. Trivedi MH, Rush AJ, Wisniewski SR, et al. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. *Am J Psychiatry*. 2006;163(1):28–40.
4. Fava M, Rush AJ, Trivedi MH, et al. Background and rationale for the sequenced treatment alternatives to relieve depression (STAR*D) study. *Psychiatr Clin North Am*. 2003;26(2):457–494, x.
5. Rush A, Fava M, Wisniewski SR, et al. Sequenced treatment alternatives to relieve depression (STAR*D): rationale and design. *Control Clin Trials*. 2004;25(1):119–142.
6. Lavori P, Rush A, Wisniewski S, et al. Strengthening clinical effectiveness trials: equipoise-stratified randomization. *Biol Psychiatry*. 2001;50(10):792–801.
7. Rush AJ. STAR*D: what have we learned? *Am J Psychiatry*. 2007;164(2):201–204.
8. Wisniewski SR, Fava M, Trivedi MH, et al. Acceptability of second-step treatments to depressed outpatients: a STAR*D report. *Am J Psychiatry*. 2007;164(5):753–760.
9. Nelson JC, Pikalov A, Berman RM. Augmentation treatment in major depressive disorder: focus on aripiprazole. *Neuropsychiatr Dis Treat*. 2008;4(5):937–948.
10. Linn BS, Linn MW, Gurel L. Cumulative illness rating scale. *J Am Geriatr Soc*. 1968;16(5):622–626.
11. 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(3):237–248.
12. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry*. 1960;23:56–62.
13. Hamilton M. Development of a rating scale for primary depressive illness. *Br J Soc Clin Psychol*. 1967;6(4):278–296.
14. Rush AJ, Bernstein IH, Trivedi MH, et al. An evaluation of the Quick Inventory of Depressive Symptomatology and the Hamilton Rating Scale for Depression: a sequenced treatment alternatives to relieve depression (STAR*D) trial report. *Biol Psychiatry*. 2006;59(6):493–501.
15. Rush AJ, Trivedi MH, Ibrahim HM, et al. The 16-Item quick inventory of depressive symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Biol Psychiatry*. 2003;54(5):573–583.
16. Gandek B, Ware JE, Aaronson NK, et al. Cross-validation of item selection and scoring for the SF-12 Health Survey in nine countries: results from the IQOLA Project. International Quality of Life Assessment. *J Clin Epidemiol*. 1998;51(11):1171–1178.
17. Endicott J, Nee J, Harrison W, et al. Quality of Life Enjoyment and Satisfaction Questionnaire: a new measure. *Psychopharmacol Bull*. 1993;29(2):321–326.
18. Wisniewski SRP, Rush AJ, Bryan CP, et al. Comparison of quality of life measures in a depressed population. *J Nerv Ment Dis*. 2007;195(3):219–225.
19. Allison P. *Missing Data. Sage University Papers Series on Quantitative Applications in the Social Sciences*. Thousand Oaks, CA: Sage; 2002.
20. Graham JW. Missing data analysis: making it work in the real world. *Annu Rev Psychol*. 2009;60(6):1–6.28.
21. Robbins AS, Chao SY, Fonseca VP. What's the relative risk? A method to directly estimate risk ratios in cohort studies of common outcomes. *Ann Epidemiol*. 2002;12(7):452–454.
22. Deakin J, O'Loughlin C. STAR*D: a summary and UK perspective. *J Psychopharmacol*. 2009;23(6):605–612.
23. Hansen RA, Moore CG, Dusetzina SB, et al. Controlling for drug dose in systematic review and meta-analysis: a case study of the effect of antidepressant dose. *Med Decis Making*. 2009;29(1):91–103.
24. Nelson JC, Papakostas GI. Atypical antipsychotic augmentation in major depressive disorder: a meta-analysis of placebo-controlled randomized trials. *Am J Psychiatry*. 2009;166(9):980–991.
25. Papakostas GI, Thase ME, Fava M, et al. Are antidepressant drugs that combine serotonergic and noradrenergic mechanisms of action more effective than the selective serotonin reuptake inhibitors in treating major depressive disorder? A meta-analysis of studies of newer agents. *Biol Psychiatry*. 2007;62(11):1217–1227.
26. Ruhé HG, Huyser J, Swinkels JA, et al. Switching antidepressants after a first selective serotonin reuptake inhibitor in major depressive disorder: a systematic review. *J Clin Psychiatry*. 2006;67(12):1836–1855.
27. Nelson JC. Augmentation and combination strategies in resistant depression. *J Clin Psychiatry*. 2009;70(6):e20.