In Review

The Sequenced Treatment Alternatives to Relieve Depression (STAR*D) Trial: A Review

Mark Sinyor, MD; Ayal Schaffer, MD, FRCPC; Anthony Levitt, MD, FRCPC

Objective: The Sequenced Treatment Alternatives to Relieve Depression (STAR*D) trial is the largest open-label, pragmatic trial that has been undertaken to examine the treatment of major depressive disorder. At a cost of US$35 million over 6 years, STAR*D sought to test the effectiveness both of pharmacotherapy and of cognitive therapy, and to ascertain whether certain treatments are more optimal after one or more failed trials.

Method: Patients (n = 2876) who presented to either a psychiatry or family practice setting seeking treatment for depression were included in the STAR*D analysis. In the 4 levels of STAR*D, patients were randomized to various treatment monotherapies, combinations, or augmentation strategies. The primary outcome was remission, based on the Hamilton Depression Rating Scale. Secondary outcomes were response, as measured by clinician and patient self-report as well as various measures of patients' level of function and (or) quality of life.

Results: Remission rates for treatment levels 1 to 2 and 3 to 4 were 18% to 30% and 7% to 25%, respectively. There was no difference in effectiveness between any treatments at any treatment level. Patients with longer index episodes, more concurrent psychiatric or general medical disorders, and (or) lower measures of baseline function were less likely to achieve remission. There were no major differences between outcomes in patients treated in primary, compared with specialist care, nor were there significant differences between depression rating scores obtained through clinician ratings, compared with self-report.

Conclusion: Results of the STAR*D trial have shed important light on the effectiveness of current treatment strategies for patients with depression.


Highlights

- In a pragmatic clinical trial setting, only a minority of depressed patients achieved remission during first-line treatment with antidepressant monotherapy.
- No specific treatment modality was statistically superior within each treatment step.
- Patients achieving remission were less likely to relapse during 1 year of naturalistic follow-up, compared with patients achieving response but not remission.

Key Words: depression, adult, pragmatic trial, antidepressant, cognitive-behavioural therapy

The STAR*D trial is one of several NIMH-sponsored effectiveness trials of mental health pharmacotherapy and psychotherapy that attempt to examine typical patients in a real-world setting. STAR*D differs from standard RCTs in 6 key ways. The first 3 pertain to study generalizability; that is, broad inclusion criteria with few exclusion criteria, the integration of patient choice, and open-label treatment. The latter 3 pertain to improving treatment strategies; that is, the use of measurement-based care, the use of remission as the primary outcome, and a sequenced treatment approach, including substitution, augmentation, and combination strategies.

Rationale for the Effectiveness Study

Commonly, RCTs in the unipolar major depression literature recruit patients who are carefully selected to have minimal psychiatric, medical, or substance comorbidities. Further, as a result of growing placebo response rates in these trials,
many investigators are choosing to further select patients for features that will make them less likely to respond to placebo, such as more severe depressive symptomatology and longer episode duration. The result is that patients in these RCTs increasingly represent only a small subset of patients routinely treated by psychiatrists and family physicians in everyday practice. In contrast, STAR*D is a so-called pragmatic trial; that is, the STAR*D trial aimed to be generalizable to clinicians in a way that standard efficacy trials cannot, by employing minimal exclusion criteria and accepting only patients seeking treatment for depression from their physician, rather than soliciting them through advertising, as is the practice in most RCTs.

A second issue regarding trial generalizability is that of patient choice. All clinicians recognize that patients’ preferences are an important consideration when deciding between multiple treatments with similar efficacy, as is the case in depression. Allowing patients to be involved in the selection of their treatments may lead to better outcomes by contributing to the therapeutic alliance and the sense of belief in, or ownership of, the treatment. However, the completely random design employed in most RCTs does not take into consideration subject choice and may, therefore, limit the generalizability of RCT results. STAR*D addresses this issue by using an equipoise-stratified randomization strategy. This approach divides patients into groups based on the treatments that they are willing or unwilling to accept, thereby permitting randomized comparisons (across groups of patients that opt for the same range of treatments) while simultaneously allowing patients to make choices based on their preferences.

The third way in which STAR*D more closely approximates the real world is by making pharmacotherapies open label. By definition, traditional double-blind RCTs keep the identity of treatments concealed from both the patient and the treating physician. Moreover, if the RCT is placebo controlled, both patient and physician are unsure as to whether the patient is even receiving an active therapy. This may have a negative impact on response and remission rates for patients participating in RCTs. STAR*D deals with this issue by making treatments known to both the patient and the treating physician, as in normal clinical practice, while simultaneously minimizing bias by keeping outcome assessors, who are independent of the treatment team, blinded to the therapy.

The latter 3 features that distinguish STAR*D from other RCTs were meant to address the need for more coherent and evidence-based treatment approaches. Standard effectiveness trials aim to test treatment strategies under real-world conditions; however, STAR*D investigators noted growing evidence that, in clinical practice, the dose and (or) duration of antidepressant treatments is often inadequate. Therefore, it was decided that STAR*D should not only examine the effectiveness of treatments in clinical practice but also attempt to optimize those treatments in a standardized, reproducible way. To that end, a measurement-based care strategy was proposed whereby symptoms and side effects would be rated at each visit and clinicians would receive automated feedback to ensure adequate and safe treatments. While such a rigorous care strategy may seem at odds with the STAR*D trial’s goal of approximating real-world conditions, particularly in primary care settings where there is significant time pressure, STAR*D investigators argue that integrating some form of management-based care into clinical practice is feasible and ought to be considered if it can be shown to improve outcomes.

The use of remission as the primary outcome of STAR*D reflects a shift by investigators toward more stringent outcome measures in an attempt to maximize the clinical utility of their results. There is a growing body of literature suggesting remission (the near-absence of depressive symptoms) is a more robust outcome than response (50% reduction in symptoms), the most common primary outcome measure in antidepressant RCTs. It has been demonstrated that patients achieving remission function better, have lower rates of suicide, and are less likely to relapse than those who achieve only response. Choosing remission, the more strict outcome criterion, by definition results in lower rates than with response as an outcome. The concern with this strategy is that it may reduce the variability of results, masking subtle differences between treatments and thus making it more difficult to demonstrate that one treatment is more effective than another. Further, in clinical practice there is no specified time in which remission must be achieved, whereas in RCTs there is an arbitrary maximum time period. Setting a time period may reduce the variability in remission rates. Nevertheless, the use of remission provides a more rigorous benchmark by which to compare treatments and STAR*D investigators also test more easily attainable outcomes such as treatment response and improved function and (or) quality of life as secondary outcome measures.

Finally, most studies in the depression literature examine single-treatment strategies, such as drug monotherapies or
combinations. Of course, in clinical practice many patients require 2 or more trials of different medications before achieving remission of their symptoms. The STAR*D investigators recognize this disparity and address it in a novel way by testing a 4-level algorithm to try to identify the optimal treatment at each step in the sequence for patients whose depression does not remit after one or more adequate trials.

Design and Methodology
STAR*D subjects were outpatients aged 18 to 75 years diagnosed with unipolar nonpsychotic MDD, and with a baseline score of 14 on the 17-item HDRS. In 2002, Zimmerman et al examined 16 RCTs published from 1994 to 2000 that used the 17-item HDRS. Among these, 14 required minimum severity scores of 18 to 22, 1 required a score of 16, and only 1 required a score of 14, as in STAR*D. The analysis showed that about 45% to 70% of depressed patients in clinical practice would be excluded from a study with a 17-item HDRS cutoff of 18 to 22, compared with less than 20% with a cutoff of 14. As such, STAR*D chose a cutoff that is lower than most efficacy trials to include the largest number of potential patients and thus maximize generalizability of results. As a consequence, STAR*D included many patients with comparatively mild MDEs.

Patients were not recruited for treatment through advertising but rather when they presented to their primary care physician or psychiatrist seeking treatment for depression. They had to have sought treatment at 1 of 23 psychiatric or 18 primary care settings across the United States participating in the study. In contrast to typical efficacy studies, STAR*D had relatively few exclusion criteria; patients who were pregnant or breastfeeding or had a primary diagnosis of bipolar, psychotic, obsessive–compulsive, or eating disorders, or had substance dependence requiring inpatient detoxification were excluded, as were those with a history of nonresponse to, intolerance of, or general medical conditions contraindicating the use of protocol medications in Levels 1 or 2 of the study. A large proportion of STAR*D participants would have been excluded from most efficacy trials in the literature. These trials would also likely have excluded the same patients that were excluded from STAR*D.

Screening was performed on 4790 patients, of whom 4177 (87.2%) consented to participate in the study. Among those patients, 1301 (31.1%) were not included because they were ineligible according to exclusion criteria (n = 136), their HDRS scores were missing or less than 14 (n = 931) or they failed to return (n = 234), leaving 2876 patients that were deemed eligible for analysis. Of note, baseline characteristics did not differ between the 2876 patients who were included in the analysis and patients who were not included. That only 136 patients, 3% of those who consented to participate, were deemed ineligible for STAR*D based on exclusion criteria belies the fact that those criteria were minimal and that the STAR*D population is far more generalizable to clinical practice than those in a typical RCT.

The diagnosis of MDD was made by treating clinicians with the aid of a checklist. Clinical research coordinators then administered the HDRS, the 16-item QIDS-C, and the 16-item QIDS-SR for patients' baseline scores of depressive symptomatology. RCTs do not typically use self-report scales as the main outcome measures; however, the QIDS-SR was used in STAR*D, in part, to test whether self-report outcome measures were comparable with those obtained via the more traditional HDRS and the QIDS-C. The primary outcome of STAR*D was remission, defined as a score of 7 on the 17-item HDRS or 5 on the QIDS-SR. Response was defined as 50% reduction in symptoms on the QIDS-SR (note that all remitters are also, by definition, responders given the requirement of a baseline score of 14 on the 17-item HDRS). HDRS scores were assessed by researchers who conducted periodic, structured, telephone-based interviews with patients. QIDS-SR scores were collected at each treatment visit. STAR*D also used an automated, telephonic, voice-response system to obtain data on subjects' functional status and quality of life.

Treatments in all levels of STAR*D were open label; that is, known to both the patient and the treating physician for the purpose of approximating real practice and to ensure rigorous dosing. However, the research outcome assessors who evaluated the primary outcome by telephone were centralized and blinded to the treatment.

STAR*D attempted to standardize and optimize treatment strategies across sites and individual clinicians. Physicians prescribing pharmacotherapy were given a manual as well as a web-based medication monitoring system for monitoring symptoms and side effects. Potential psychotherapists were given readings and a 2-day workshop on a standard approach to CBT. Psychotherapists then had to demonstrate competence in treating one patient meeting STAR*D eligibility criteria. As with subjects on pharmacotherapy, symptoms were monitored via a web-based system.

The STAR*D design included 4 different treatment levels. In Level 1, all patients were treated with flexible doses of citalopram for up to 14 weeks. Patients achieving remission or response after an adequate trial were allowed to enter a 12-month naturalistic follow-up. However, patients who achieved response but not remission were strongly encouraged to enter the second treatment level along with the nonresponders and those who could not tolerate citalopram owing to side effects.

Level 2 consisted of 7 different treatment options including 3 augmentation strategies (citalopram + bupropion SR, citalopram + bupropion, citalopram + CBT) and 4 switch strategies (bupropion SR, sertraline, venlafaxine XR, CBT). In Level 2, subjects received pharmacotherapy and (or) psychotherapy for 12 weeks. CBT entailed 16 sessions, with biweekly sessions for the first 4 weeks and weekly sessions for the final 8 weeks. As in Level 1, remitters and responders were offered naturalistic follow-up and those with unsatisfactory response (intolerance or lack of
remission) were encouraged to enter Level 3.\(^{20,23}\) In another
effort to approximate real-world conditions, Level 2 intro-
duced the element of patient choice via “equipoise-stratified
randomization.”\(^{20}\) That is, subjects could opt not to
receive certain treatment options according to their prefer-
ences (in consultation with their treating physicians). In this
scheme, they choose only to be randomized to one of the
remaining treatment options that they are comfortable receiv-
ing.\(^{20}\) For example, subjects could choose to be randomized to
any switch strategy but no augmentation strategies; or, alter-
natively, they could choose any medication but not CBT.\(^{20}\)

STAR*D investigators wanted all subjects eligible for Level 3
to have failed 2 adequate trials of pharmacotherapy for a
reasonable comparison.\(^{25}\) To that end, Level 2A was created
so that subjects with unsatisfactory response that had been
augmented with or switched to CBT (that is, had only received
citalopram as pharmacotherapy) would be randomized to be
switched to either bupropion SR or venlafaxine XR using the
same protocol from Level 2.\(^{16,23}\)

Level 3 was again structured using equipoise-stratified
randomization, with treatment options being an augmentation
strategy (lithium or T3), a switch strategy (mirtazapine or
nortriptyline) or to either strategy for 12 weeks.\(^{23,24}\) Subjects
randomized to an augmentation strategy received lithium or
T3 in addition to ongoing treatment with citalopram, bupropion
SR, sertraline or venlafaxine XR. If bupropion SR or bupirone augmentation was used in Level 2, these
comedications were discontinued.\(^{23}\) Finally, subjects who failed
Level 3 could be entered into Level 4 in which they were ran-
domized to treatment with either tranylcypromine or the com-
bination of venlafaxine XR and mirtazapine.\(^{25}\)

Patients who achieved remission (or, less preferably, response)
during any treatment level and who consented to 12-month naturalistic follow-up were assessed at treatment
visits every 2 months.\(^{26}\) Though patients were encouraged to
continue the same medication on which they achieved remis-
sion or response and at the previously effective dosage,
patients were allowed to change dosages, medications, or par-
ticipate in any psychotherapy modality prescribed in consul-
tation with their treating physicians.\(^{26}\)

Funding
Funding for STAR*D was provided using federal funds from
the NIMH at a total cost of US$35 million during 6 years.\(^{27}\)
While the trial did not receive direct funding from pharmaceu-
tical companies, all study medications not available in generic
preparations (all but lithium, nortriptyline, and
tranylcypromine) were provided at no cost to patients by the
various patent-holder companies.\(^{27}\) Pfizer also provided
Viagra at no cost to patients experiencing sexual side effects
to medication.\(^{27}\) STAR*D did not reimburse patients random-
ized to cognitive therapy for copayment charges for
insurance-covered sessions; that is, patients provided some of
their own funding for psychotherapy.\(^{20}\)

Principal Outcomes and Their Implications
for Clinical Practice and Research
Remission and Response Rates
HDRS remission rates were 28% using intent-to-treat analy-
sis in Level 1 and ranged from 18% (sertraline switch) to 30%
(bupropion augmentation) in Level 2 (Figure 1).\(^{5,20,28,29}\)
HDRS remission rates ranged from 12% (mirtazapine
switch) to 25% (T3 augmentation) and 7% (tranylcypromine
switch) to 14% (venlafaxine XR + mirtazapine switch) in
Levels 3 and 4, respectively (Figure 1).\(^{23–25}\) QIDS-SR remis-
sion rates, which were generally similar or slightly higher
than the HDRS rates, and response rates are shown in
Figure 2.\(^{26}\) Rates of remission and response were higher in
Levels 1 and 2 than in the latter 2 levels. Importantly, no sta-
tistically significant differences in remission rates were
found when comparing different treatments in the same treat-
ment level.\(^{30}\) Cumulative QIDS-SR remission rates for study
Levels 1 to 4 were 37%, 56%, 62%, and 67% for patients who
remained in the study at each of the respective levels (Figure 3).

There are 2 key conclusions that can be drawn from STAR*D
remission and response data. The first is that many people
who do not achieve remission or response within several
weeks of the initiation of a first therapy may do so by the end
of 14 weeks and even more will remit or respond after a sec-
ond treatment step. The second key conclusion is that the
chances of achieving good outcomes fall dramatically after
the first 2 treatment steps. Therefore, questions remain
regarding how clinicians should initiate and continue treat-
ment for depression. The potential risks and benefits of more
aggressive dosing of monotherapy agents, opting for an aug-
mentation or combination strategy earlier in the course of
treatment, and extending the duration of antidepressant trials
all require further evaluation.

The finding that there were no statistically significant differ-
ences in remission rates between treatments in the same level
is also noteworthy, although there were numerical differ-
ences between some treatments. Numerous factors complicate
this result, including heterogeneity of the study
population, variability within the data, and study power.
Power is a particularly important consideration in STAR*D.
To show statistical significance, one would ideally have
wanted Levels 2 to 4 to be well powered. Power was an issue
owing to multiple treatment comparisons in Levels 2 and 3
that, as a result of the equipoise-stratified randomization
strategy, were further complicated by multiple
randomizations based on the preferences of particular
cohorts of patients. Large decreases in the number of patients
in each successive level also negatively affected study
power. One might speculate whether statistically significant
differences might have been found had STAR*D, particu-
larly Levels 3 and 4, been powered more optimally.

Recent work by Sinyor et al\(^{13}\) examined 91 RCT studies (88
published and 3 unpublished) looking at drug monotherapies
for MDEs, studies that were similar in duration to Level 1 of STAR*D, although they were double-blinded. Fifty-two of those studies quoted intent-to-treat remission rates (37 with the same criterion as STAR*D—a score of 7 on the HDRS) with an average remission rate of 44%, compared with 28% in Level 1 of STAR*D. Response rates were also much higher in our review of RCTs (62%) than in Level 1 of STAR*D (49%). A subanalysis of RCTs that involved patients randomized to citalopram, the therapy used in Level 1, yields the same comparatively high remission (44%) and response rates (65%). It may be that this disparity occurred because STAR*D included patients who, a priori, owing to medical and (or) psychiatric comorbidities or other factors (including severity of depression or demographic differences), were less likely to achieve remission or response. However, there may be some other systematic issue in the STAR*D design that mitigated remission rates or some design feature of other RCTs that artificially elevated remission rates, in some, as yet, unknown fashion.

Secondary Outcomes and Their Implications for Clinical Practice and Research

Demographic Features and Features Associated With Remission

Patients enrolled in Level 1 of STAR*D had a mean age of 40.8 years and reported a median length of current MDE of 7.8 months, with 25% reporting chronic depression (greater
than 2 years). Among Level 1 patients, 63.7% were female, 75.8% were Caucasian, 41.7% were married, 56.2% were employed, and 65.2% had at least one other comorbid psychiatric disorder. Patients’ mean 17-item HDRS score on study entry was 21.8. Patients had a mean of 5.9 prior MDEs, and 53% were considered to have anxious depression where this was defined as an HDRS anxiety or somatization factor score of 7. Patients were more likely to achieve remission if they were Caucasian, female, employed, or had higher levels of education or income. Patients were less likely to achieve remission if they had longer index episodes, more concurrent psychiatric disorders (especially anxiety disorders and drug abuse), more general medical disorders, had lower measures of baseline function or quality of life, and lacked private insurance. Future reports will probe whether subgroups of patients benefited more from particular treatments or treatment sequences. This information would be highly clinically relevant given that there is little evidence to guide clinicians around the choice of particular treatments for different patient populations.

**Longer-Term Prognostic Features**

Remitters or responders who consented were entered into 12-month naturalistic follow-up. Patients in remission at

---

**Figure 2 QIDS-SR remission and response rates across all 4 treatment levels**

<table>
<thead>
<tr>
<th>Treatment Level</th>
<th>Total Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td>3671</td>
</tr>
<tr>
<td>Level 2</td>
<td>1439</td>
</tr>
<tr>
<td>Level 3</td>
<td>390</td>
</tr>
<tr>
<td>Level 4</td>
<td>123</td>
</tr>
</tbody>
</table>

Differences in remission rates across a particular treatment level were not statistically significant at any of the 4 levels. The definition of response for Levels 2 to 4 was ≥50% improvement from QIDS-SR score at entry to that level, not at entry to Level 1.
follow-up entry were less likely to relapse than those who had only achieved response (33.5% and 58.6%, respectively, for Level 1). More acute treatment trials were associated with higher relapse rates; for example, relapse rates for remitters and nonremitters were 50.0% and 83.3%, respectively, for those who entered follow-up after Level 4, in contrast to the lower rates of relapse listed above for those who entered follow-up after Level 1. These findings support previous data that found remission to be associated with improved long-term outcomes.

Primary Care, Compared With Psychiatric Care, and the Use of Measurement-Based Care

Patients who presented to primary care settings were generally similar in clinical presentation to those who presented to psychiatric care facilities, with the exception of having higher rates of general medical conditions and lower rates of past suicide attempts and comorbid psychiatric disorders. There was no difference between remission and response rates or time to remission and (or) response between primary and psychiatric patients in Level 1 of STAR*D. This lack of difference may be due to the higher quality of measurement-based care. Analysis of Level 1 clinical encounters demonstrated 85% fidelity between treatments and measurement-based care recommendations. Investigators describe most deviations from recommendations as justifiable and call attention to the statistically identical outcomes between primary care and psychiatric settings as evidence that measurement-based care is effective at standardizing and optimizing treatments. These findings highlight the challenges faced by family physicians that treat depressed patients. These data suggest that patient severity may be similar, yet the tools available to primary care physicians (for example, time and training) may not be sufficient. STAR*D results argue for improved methods of knowledge translation and (or) continuing education for family physicians. They also demonstrate the need for the development and use of appropriate, efficient tools to enhance management of depression in primary care.

Treatment Duration and Time to Remission

The ideal duration for an antidepressant trial is not well characterized. The traditional 4- to 6-week time frame may not be most appropriate given that a subset of patients with minimal improvement after that period go on to achieve remission by week 12. In a fluoxetine trial, Quitkin et al found that 31% to 41% of nonresponders and 48% of partial responders at week 6 had remitted by week 12. STAR*D results suggest that a medication trial should be continued if there is even a modest (20%) reduction in symptoms at 6 weeks. Indeed, among remitters, the mean time to remission was 6.3, 5.4, 5.6, and 7.4 weeks for Levels 1 to 4, respectively, suggesting that longer trials of adequately dosed pharmacotherapies may be warranted both in clinical practice as well as in research studies testing antidepressant therapies.

Choosing Between Treatments That Are Equally Effective

Although there were no statistically significant differences in remission or response rates between different treatments at each treatment level, certain treatment options had advantages over others in terms of side effect profile or time to remission or response. CBT, while equally effective compared with medication switch and augmentation strategies tested in Level 2, was associated with fewer side effects than pharmacotherapy.
For example, 34% of patients randomized to a Level 2 medication switch reported at least a moderate level of side effect burden, compared with none who were randomized to CBT, although this did not lead to a significant difference in the number of dropouts between groups owing to side effects. Nonetheless, CBT may be more acceptable to patients who are concerned about medication side effects. CBT was also associated with longer times to remission than Level 2 pharmacotherapies (55 days, compared with 40 days, \( P = 0.02 \)) and may be a less desirable option if speed of improvement is of critical importance. Nevertheless, the comparison between CBT and pharmacotherapy is imperfect in that medications were taken daily while CBT sessions were biweekly or weekly.

T3 augmentation is another example of a treatment that performed better on secondary outcome measures. Patients in Level 3 of STAR*D who were randomized to T3 augmentation reported fewer side effects and were less likely to drop out owing to side effects than those who received lithium augmentation. While a strong body of evidence for lithium augmentation suggests an ongoing role for this treatment modality, short- and long-term tolerability concerns remain an obstacle.

**Subject Attrition**

Attrition was a major issue in STAR*D, with 26% of enrolled patients dropping out of the study for nonmedical reasons. About one-third of dropouts did so after only the baseline visit and were considered “immediate attrition subjects,” while those who dropped out after at least one postbaseline visit were considered “later attrition subjects.” Both types of attrition were associated with younger age and lower levels of education. Immediate attrition was also associated with higher perceived levels of mental health functioning, while later attrition was associated with being African American. Patients who had experienced more than one MDE were less likely to drop out. It has been argued that patients with features associated with higher attrition rates might benefit from more aggressive treatment strategies that might yield a quicker response.

A recent meta-analysis of double-blind RCTs studying depression reported attrition rates on the order of 10%, much lower than the attrition rate of 26% in STAR*D. This relatively high attrition may have occurred owing to STAR*D’s relatively nonselected patients, significantly longer duration for those who continued to Levels 2 to 4, and (or) the relative inexperience of primary care and CBT providers.

**Patient Choice, Equipoise-Stratified Randomization**

One of the most noteworthy findings of STAR*D was that patient preference was a significant factor in treatment decisions; only 1.5% of patients (21/1439) enrolled in Level 2 were willing to accept to be randomized to all 7 possible augmentation or switch options. A plurality of Level 2 (70%, combined) accepted only a medication switch (583/1439) or medication augmentation (430/1439). Far fewer STAR*D participants would accept CBT as an option than was expected or than had been willing to accept CBT in prior studies done by the same investigators. Had all patients been randomized to all treatments, 411 would have been randomized to CBT; however, by accounting for patient preference, only 147 were randomized to CBT. This low acceptance rate may be because patients who were more interested in psychotherapy may have been less likely to enter a study that required a trial of pharmacotherapy in Level 1, or it may be that patients were discouraged because they might have had to travel to different treatment sites for CBT or spend on copayments. One could argue that future studies should try to eliminate these disincentives. Concurrently, this demonstrates how real-world considerations influence patients’ choices in ways that are not tested in standard RCTs.

Although Canada has a publicly funded health care system, the barriers to patients getting CBT are not insignificant, with data suggesting that access to this kind of therapy is scarce across Canada. If a patient cannot access a treatment, the theoretical cost to the patient of that treatment is irrelevant. However, if access could be improved, the publicly funded system would offer significant advantages over the copayment strategy that might have negatively influenced CBT results in STAR*D.

**Self-Reporting Scales as Primary Outcome Measures**

An analysis by STAR*D investigators found that the results obtained using the QIDS-SR were in agreement with those obtained via the clinician-rated QIDS and the HDRS. It further showed that the self-rating scale was just as sensitive in detecting symptom change, remission, and response as the clinician rating scales. These findings suggest that patient self-report measures may be helpful tools for evaluating outcomes in RCTs for depression, and indeed may be useful to clinicians in evaluating outcomes to treatment in everyday practice. A full discussion of use of self-report measures in clinical practice is beyond the scope of this review.

**Other Considerations**

The undertaking of a multicentre, US$35-million trial is impressive and daunting in that it necessitates choices that both expand and limit the clinical utility of findings. For example, the STAR*D design required a single antidepressant monotherapy in Level 1 and investigators opted to choose citalopram. One might wonder whether different results would have been found with another agent. Although there is no strong evidence that certain antidepressants are better than others for achieving remission or response in the efficacy literature, it is not clear whether all antidepressants are equivalent in pragmatic trials such as STAR*D. Indeed, designing an option for different treatment strategies at Level 1 might have enhanced the generalizability of Level 1 results, although this approach would obviously have made Level 2 augmentation or switch randomization strategies more complicated. It would also have been interesting to see what the placebo remission or response would have been had it been
practical to include a placebo group, particularly because it is more difficult to demonstrate drug-placebo differences when patients with relatively low baseline depression scores are included, as in STAR*D. Naturally, the inclusion of a placebo arm would have meant a departure from the principle of pragmatism of the trial. While the approach taken by the STAR*D investigators did permit the inclusion of a broad range of patients, it may have resulted in 2 other problems. First, allowing low baseline depression scores makes it easier to achieve remission. Second, allowing lower scores would make it difficult to demonstrate not only statistical separation between a drug and a placebo but also such a difference between different drugs. This tendency toward equivalence in milder depression may be part of the reason that no treatments were found to be more effective than others.

STAR*D investigators put considerable effort into including CBT arms into what was essentially a pharmacotherapy effectiveness trial. However, only one psychotherapy modality was tested and there were at least 2 disadvantages to those patients randomized to CBT. First, the CBT group entered Level 2 with higher scores of functional impairment and lower quality of life scores than those who were assigned to pharmacotherapies\(^1\) and, second, STAR*D employed relatively inexperienced psychotherapists, and previous research has demonstrated that therapist experience may significantly impact on outcomes.\(^20,44–46\)

**Conclusion**

Results of the STAR*D trial have shed important light on the effectiveness of current treatment strategies for patients with depression. STAR*D encompasses an enormous amount of information with several key findings. Many patients who did not achieve remission or response after several weeks of the first antidepressant did achieve these outcomes by the end of 14 weeks. Another large group who did not remit or respond after the first antidepressant did so after a second antidepressant trial. Positive outcomes decreased precipitously in subsequent antidepressant trials (3 and 4). Further, patients who achieved remission, rather than response, were far less likely to have relapsed at 12-month naturalistic follow-up, suggesting that remission may be a more robust outcome. Finally, STAR*D investigators found that their measurement-based care strategy was effective at standardizing and optimizing treatments, which may provide avenues for improving future treatments both in the primary care and in the psychiatric setting.

The enhanced generalizability to real-world patients justifies the substantial investment in STAR*D and the broad interest in the results of this study. While the primary outcome measures did not significantly differ between treatments, some of this may be attributed to methodological issues outlined in this review. Large pragmatic trials such as STAR*D may, in the end, produce more questions than answers, yet future studies can build on the important information that has been obtained.

**Acknowledgements**

Dr Sinyor has no conflicts to report. Dr Schaffer has in the past received speakers’ bureau honoraria, advisory panel funding, and (or) research grants from Eli Lilly Canada, AstraZeneca Canada, Lundbeck, Biovail, Sanofi-Aventis, and Servier Canada. Dr Levitt has acted as a consultant for Janssen Ortho, Biovail Corp, and Eli Lilly Canada. The Canadian Psychiatric Association proudly supports the In Review series by providing an honorarium to the authors.

**References**

Résumé : Revue de l’étude de traitements séquences thérapeutiques de la dépression (STAR*D)

Objectif : L’étude de séquences thérapeutiques possibles pour traiter la dépression (STAR*D) est l’essai pragmatique le plus vaste qui ait été entrepris pour examiner le traitement du trouble dépressif majeur. Au coût de 35 millions de dollars US sur 6 ans, STAR*D a cherché à vérifier l’efficacité à la fois de la pharmacothérapie et de la thérapie cognitive, et à évaluer si certains traitements sont plus optimaux après un ou plusieurs essais ratés.

Méthode : Des patients (n = 2876) qui se sont présentés soit chez un psychiatre soit médecin de famille à la recherche d’un traitement pour la dépression ont été inclus dans l’analyse STAR*D. Dans les 4 niveaux de STAR*D, les patients ont été affectés au hasard à diverses monothérapies, combinaisons ou stratégies d’augmentation de traitement. Le résultat primaire était la rémission, d’après l’échelle de dépression de Hamilton. Les résultats secondaires étaient la réponse, telle que mesurée par le clinicien et l’autodéclaration du patient, et par diverses mesures du niveau de fonctionnement et (ou) de la qualité de vie des patients.

Résultats : Les taux de rémission pour les niveaux de traitement 1 à 2 et 3 à 4 étaient de 18 % à 30 % et de 7 % à 25 %, respectivement. Il n’y avait aucune différence d’efficacité entre tous les traitements, à tout niveau de traitement. Les patients ayant des épisodes de référence plus longs, plus d’autres troubles psychiatriques ou médicaux généraux co-occurrents, et (ou) des mesures plus faibles de fonctionnement au départ étaient moins susceptibles d’obtenir une rémission. Il n’y avait pas de différences majeures entre les patients traités dans des soins de première ligne, comparativement aux soins spécialisés, ni de différences significatives entre les scores d’évaluation de la dépression obtenus par les cliniciens, comparativement à l’autodéclaration.

Conclusion : Les résultats de l’étude STAR*D éclairent l’efficacité des stratégies de traitement actuelles pour les patients souffrant de dépression.