Sequenced Treatment Alternatives to Relieve Depression

(STAR*D)

Revised June 28, 2002

Table of Contents

Table of Figures

Table of Tables

Abstract: STAR*D

Public Health Significance: STAR*D focuses on major depressive disorder (MDD), which is a common, usually recurrent, and often chronic disorder. MDD causes substantial disability and costs the U.S. over $40 billion/year (both direct and indirect costs).[75] Many treatments (medication, psychotherapy, and electroconvulsive therapy [ECT]) for depression have efficacy established by randomized controlled trials (RCTs).[61] However, the actual acceptability, clinical benefit, and side effect burden of these treatments in populations seen in representative clinical settings, is not well known. It is also unclear how to help depressed participants who respond but who do not remit with an initial antidepressant treatment. Symptomatic responses that fall short of complete symptomatic remission (sometimes termed partial response or partial remission) are frequent,[52] are associated with continuing disability, and have a poorer prognosis than when complete remission is attained.

Objectives: STAR*D is prospectively designed to determine the comparative EFFECTIVENESS of different treatment options for participants with MDD. It will evaluate the comparative effectiveness of these treatments when they are used to either augment the previous treatment or as new treatments for participants without a satisfactory response to an initial selective serotonin reuptake inhibitor (SSRI) medication. Assignments to treatment options are randomized at all levels of treatment (2, 2A, 3, 4) after initial treatment with an SSRI. Clinical outcomes include symptoms, function, side effect burden, quality of life, and participant satisfaction. These outcomes are evaluated by independent assessments masked to treatment assignments.

STAR*D results will provide an empirical basis for practice guidelines, thereby reducing the reliance on clinical consensus. Results will also provide descriptive information about selected treatment tactics (e.g., at what dose and for how long should a treatment be conducted before declaring failure and, therefore, moving to the next level of treatment?).

STAR*D also has the following secondary objectives. For those who respond satisfactorily to any treatment and who enter the 12-month follow-up, STAR*D will determine the incidence, nature, and timing of relapses (< 6 months after remission) or recurrences (> 6 months after remission). STAR*D will also relate the clinical outcomes to the treatment costs and cost-offsets entailed in following the multistep treatment sequence.

Levels of Treatment: The STAR*D project will enroll 4,000 outpatients (ages 18-75) with nonpsychotic MDD into the treatment protocol and initially treat them with citalopram (CIT) (the Level 1 treatment) for a minimum of 8 weeks, with strong encouragement to complete 12 weeks in order to maximize benefit. All participants will also receive a brief depression educational program.

From the 4,000 participants who enter the treatment protocol, we expect 2,000 participants to not have a satisfactory therapeutic response to CIT. These 2,000 individuals are eligible for 7 different treatment options at Level 2. These options may be

conceptualized as representing two overall treatment strategies: 1) Medication or Psychotherapy Switch — switching from CIT to another antidepressant medication or Cognitive Psychotherapy (CT), and 2) Medication or Psychotherapy Augmentation — augmenting CIT with a second medication or CT. Within Level 2, multiple sets of treatment options (substrategies) will be compared, reflecting competing clinical rationales as to the types of medications to use for participants without a satisfactory response to an initial SSRI.

For those who switch treatments at Level 2, sertraline (SER) (a second SSRI), venlafaxine (VEN) (an antidepressant with both noradrenergic and serotonergic effects), bupropion (BUP) (an antidepressant with both noradrenergic and dopaminergic effects), or CT are available. Likewise, within the Medication or Psychotherapy Augmentation strategy, the three treatments for augmenting CIT are BUP, buspirone (BUS) (an anti-anxiety medication), or CT.

At treatment Levels 2, 2A, 3, and 4, participants will be randomly assigned to all treatment groups and options found acceptable to them. For example, at Level 2, before randomization, participants will indicate whether they wish to continue or discontinue the CIT from Level 1. They will be encouraged to accept both possibilities, but some may decide they do not wish to continue, while others may insist only on continuing CIT. This first decision determines if participants are eligible for either the Medication or Psychotherapy Switch or Augmentation groups, or both. Secondly, participants will be asked if they are willing to accept CT as a treatment option at Level 2. Those declining CT altogether would be assigned to either a Medication Switch or a Medication Augmentation group. While all participants eligible for Level 2 will be asked to accept multiple strategies, they may continue in the protocol so long as a randomized assignment can be made among treatments found acceptable. Participants cannot select among specific medication treatment options (i.e., they must accept both of the medication augmentation options or all three medication switch options). Participants who will not accept randomization to any option will exit the study.

When data are analyzed for Level 2, the substrategies are defined by participant acceptability: 1) switching to a new treatment (Medication or Psychotherapy Switch); 2) augmenting with a second treatment (Medication or Psychotherapy Augmentation); 3) switching, but only to a new medication (Medication Switch); 4) augmenting, but only with medication (Medication Augmentation); 5) switching or augmenting, but only with medication (Medication Augmentation or Switch); or 6) switching to or augmenting with only cognitive therapy (Psychotherapy Augmentation or Psychotherapy Switch).

For only those participants without a satisfactory response to CT (used either as a switching or augmenting option) at Level 2, additional medication treatments are provided (randomized) at what is called Level 2A. Assignment to one of two Medication Switch options (VEN or BUP) ensures that all participants who enter Level 3 will have not had a satisfactory response to two different medications before entering Level 3.

All participants who do not respond satisfactorily at Level 2 or 2A are eligible for Level 3. Level 3 includes medication switch to either mirtazapine (MIRT) or nortriptyline (NTP), a tricyclic antidepressant (TCA), and medication augmentation with either lithium (Li) or thyroid hormone (THY) (Figure 4).

Level 4 treatments include a medication switch to either tranylcypromine (TCP), a monoamine oxidase inhibitor (MAOI), or to a combination of mirtazapine and venlafaxine (MIRT+VEN), using random assignment for all without a satisfactory response to Level 3.

In summary, the treatment levels consist of:

Level 1: Citalopram

Level 2 (with randomized assignment to 7 treatment options within as many of the above six substrategy groups deemed acceptable and medically safe):

— Medication or Psychotherapy Switch strategy options: switch to a) SER, b) VEN, c) BUP, or d) CT.

— Medication or Psychotherapy Augmentation strategy options: add to CIT either a) BUS, b) BUP, or c) CT.

— Medication Only Switch or Medication Only Augmentation substrategies are available for participants for whom CT is unacceptable.

Level 2A (with randomized assignment to treatment options for participants who have not responded adequately to cognitive therapy as a switch or augmentation strategy at Level 2):

— Medication Switch strategy options: switch to a) VEN or b) BUP.

Level 3 (with random assignment to treatment options deemed acceptable and medically safe):

— Medication Switch strategy options: switch to either a) MIRT or b) NTP.

— Medication Augmentation strategy options: add (to Level 2 or 2A psychopharmacological treatment) either: a) lithium or b) thyroid hormone (THY).

Level 4 (with random assignment to treatment options deemed acceptable and medically safe):

— Medication Switch options: switch to either a) TCP or b) MIRT+VEN.

Follow-up Protocol: Participants who respond satisfactorily to any treatment are followed naturalistically.

In summary, STAR*D plans to test sequenced treatments, organized in a series of steps according to the algorithm outlined above. Levels 2, 2A, 3, and 4 involve randomized comparisons between the treatment options for those who have not remitted at the previous level.

Clinic Visits: During all clinic visits in all treatment levels and in follow-up, symptomatic status will be measured by the 16-item Quick Inventory of Depressive Symptomatology – Clinician-Rated (QIDS-C_{16}).[196] In addition, the Clinical Record Form (CRF) will record the Clinical Global Impression-Improvement (CGI-I),[79] the Frequency and Intensity of Side Effects Rating (FISER), the Global Rating of Side Effect Burden (GRSEB) ratings, and the Patient Rated Inventory of Side Effects (PRISE). The clinic ratings are used by the clinician to guide the implementation of the protocol.

Research Outcomes: Research outcomes are collected independent of clinic visits by persons not engaged in the treatment process. Research outcomes are collected at pretreatment and exit from each treatment level and at months 3, 6, 9, and 12 during post-level follow-up.

Research outcomes include the following domains and measures: symptoms: (the 17-item Hamilton Rating Scale for Depression (HAM-D_{17}),[80,81,179,241] the 30-item Inventory of Depressive Symptomatology - Clinician-Rated (IDS-C_{30}),[198,199] and the QIDS$_{16}$; function: 12-item Short-Form Health Survey (SF-12),[232] the 6-item Work Productivity and Activity Impairment Questionnaire (WPAI),[186] and the 5-item Work and Social Adjustment Scale (WSAS);[165] quality of life: the 16-item Quality of Life Enjoyment and Satisfaction Questionnaire (Q-LES-Q);[45] side-effects: the 2-item FISER, and the 1-item GRSEB; participant satisfaction: the 2-item Likert-type Patient Satisfaction Inventory (PSI); and health care utilization gathered by a modified Utilization and Cost Patient Questionnaire (UAC-PQ_{15}).[25,36,108,114,139,183]

Research outcomes are collected by telephone interview with Research Outcomes Assessors (ROAs) masked to treatment (HAM-D_{17} and IDS-C_{30}) and by an interactive voice response (IVR) system (QIDS$_{16}$, WPAI, WSAS, PSI*, Q-LES-Q, SF-12, UAC-PQ_{15}, FISER*, GRSEB*) using push-button telephones. The IVR will also collect the QIDS$_{16}$, the WPAI, WSAS, and FISER/GRSEB at week 6 in each treatment level to characterize the early trajectory of response and utilization information for all participants. The identical interim outcomes will be obtained monthly on participants in post-level follow-up at months 1, 2, 4, 5, 7, 8, 10, and 11 (i.e., at months when the full research outcomes battery is not collected).

Analytic Plans: To account for differences in participants' acceptance of the various treatment options, we have created for Levels 2, 3, and 4 specific "acceptability strata."

---

* PSI, FISER, and GRSEB are obtained only at post-baseline assessments.

Participants are stratified according to their patterns of acceptance of the 7 treatment options at Level 2, of the 4 treatment options at Level 3, and randomized among whatever treatment options are included in the strategies they accept. In analyses comparing particular sets of treatment options (i.e., treatment substrategies) at Levels 2 and 3, the outcome data will be pooled across all participants assigned to those treatments under conditions of potential randomization to any of them. Information specifying these participants' varying acceptability strata provides a measure of their treatment preferences and will be used in analyses as a covariate to control for possible effects of the preference variations.

In general, in analyzing the study results for Level 2, the plan is to conduct cross-substrategy comparisons for the following subsets of treatment options: 1) All strategies (possible only for participants who accept to be randomized to all treatment options, or "universal donors"); 2) Medication or Psychotherapy Switch; 3) Medication or Psychotherapy Augmentation; 4) Medication Augmentation or Medication Switch; 5) Psychotherapy Augmentation or Psychotherapy Switch; 6) Medication Switch; and 7) Medication Augmentation.

Infrastructure: The STAR*D infrastructure includes the National Coordinating Center (NCC) in Dallas, three Special Function Regional Centers (RCs) (Boston, New York, and Pittsburgh), and RCs at Ann Arbor, MI; Chapel Hill, NC; Chicago, IL; Dallas, TX; Los Angeles, CA; Nashville, TN; Richmond, VA; San Diego, CA; Tulsa OK: Tuscaloosa, AL; and Wichita, KS. The NCC is responsible for all RCs. Each RC is responsible for the recruitment, retention, and safety of their study participants and for the acquisition of the study data from at least two but not more than three specific clinical sites (physically distinct locations where participants are treated). Clinical sites will include both public and private sector practices that provide primary or specialty care, to ensure a broad generalizability of findings. The STAR*D infrastructure also provides opportunities to conduct ancillary studies consistent with improving the care of persons with treatment-resistant depression, that otherwise do not impede the conduct of the primary protocol.

I. Specific Aims/Objectives

STAR*D has several main objectives and is powered to assess the effectiveness of a sequence of treatments at various levels of treatment. Most of these objectives entail evaluating participants at successive points in a sequence of treatments, with progression to each subsequent treatment depending on the adequacy of the participants' response to the preceding treatment:

(1) To determine the relative effectiveness of seven different treatment options at Level 2, following lack of symptom remission with an adequate trial (in terms of both dose and duration) with a selective serotonin reuptake inhibitor (SSRI) (citalopram) (CIT). Treatments include venlafaxine (VEN), sertraline (SER), bupropion (BUP) and cognitive therapy (CT) as monotherapies, or CT, BUP, or buspirone (BUS) used as augmenting agents.

(2) For those without remission after two adequate trials with antidepressants, to determine the relative effectiveness of four treatment options at Level 3: mirtazapine (MIRT) or nortriptyline (NTP) as monotherapies, or lithium (Li) or thyroid hormone (THY) used as augmenting agents.

(3) For those without remission after Levels 1, 2, and 3, to determine the relative effectiveness of two treatment options at Level 4: switching to tranylcypromine (TCP) or to the combination of mirtazapine and venlafaxine (MIRT+VEN).

(4) To characterize the course of illness after adequate response to and continuation on the treatments found effective for individual participants.

(5) Across all levels of treatment, to determine the relative health care utilization and costs for those who receive protocol-based treatment.

(6) To compare direct care costs with outcomes across different treatment options and across provider and payer venues.

The specific aims for each treatment level and for health care utilization and cost are considered separately in the following sections A-F:

A. Participants Who Enter Level 1 Treatment (Initial treatment) (See Figure 1, page 23)

A-1.   Primary Objective: (Effectiveness)
(1) Determine the time course of clinical response and symptomatic remission and other clinical benefits (symptoms, function, attrition, side effect burden, and satisfaction) to CIT as the initial treatment for clinical depression.

A-2.   Secondary Objectives: (**Predictors**)

(1) Identify baseline clinical and demographic predictors of response and remission.

(2) Identify baseline predictors of earlier versus later response and remission.

(3) Evaluate the predictive value of early symptom change (e.g., baseline to week 4) or in identifying those who will respond (or remit) later (e.g., by week 12).

(4) Identify baseline predictors of intolerance to CIT.

B. Participants Randomized to Level 2 Treatment Options (Participants Without a Satisfactory Response to CIT) (See Figure 2, page 24)

B-1.   Primary Objective: (**Effectiveness**)

(1) Determine the most effective seven treatment option(s) clustered within the six substrategy groups for individuals eligible for Level 2 treatments. The best treatment options within each substrategy group will be identified. Then, the best treatment options will be compared across substrategy groups. Direct comparisons will only be made among participants who agree to randomization within and across substrategy groups.

B-2.   Secondary Objectives: (**Acceptability and Predictors**)

Exploratory analyses will be conducted to identify predictors of: acceptability of the various treatment options and substrategy groups, overall response to different substrategies, and response to specific treatment options. These predictor analyses will be done within and across substrategies. The following are illustrative examples.

(1) Determine which treatment options or substrategy groups (Medication and Psychotherapy Switch, Medication and Psychotherapy Augmentation, Medication Switch, Medication Augmentation, Medication Augmentation or Switch, Psychotherapy Augmentation or Psychotherapy Switch) are most acceptable.

(2) Assess the relationship between the degree of symptomatic improvement and side effect burden with Level 1 treatment, and the acceptability and/or response to Level 2 substrategies. Some participants will benefit modestly from the initial SSRI, while others will have substantial improvement but continue to have residual

symptoms.* We will determine whether the degree of symptomatic or functional improvement with CIT affects participants' acceptance of and response to the six treatment substrategy groups offered at Level 2. We expect that those who experience less symptomatic benefit (or intolerance) with CIT will be more likely to decline augmentation (whether psychotherapy or medication) and be more likely to accept a switching substrategy than those experiencing greater benefits with CIT.

(3) Determine if there are pretreatment (baseline) predictors of acceptability or response to the six treatment substrategies in Level 2. Exploratory analyses will evaluate the effect of specific baseline variables including PARTICIPANT FEATURES (e.g., age, gender, ethnicity, socioeconomic status, etc.), ILLNESS FEATURES (e.g., prior history of treatment, family history of mood disorders, atypical or melancholic symptom features, depressive symptom severity, presence or absence of concurrent general medical or psychiatric conditions, etc.), and CARE FEATURES (e.g., clinician type, clinician experience; primary versus specialty care settings) on the acceptability of the six treatment substrategies and on the comparative clinical benefits of the several different treatment options.*

(4) Determine whether there are predictors of response to each treatment option, as well as differences in predictors between treatments by examining the relevance of the PARTICIPANT FEATURES, ILLNESS FEATURES, and CARE FEATURES (see above) for clinical outcomes. For example, we will assess the effect of prior psychotherapy during the current episode on the rate of response to cognitive therapy.

These exploratory analyses will:
- Determine the impact of concomitant general medical conditions (GMCs) on key clinical outcomes (i.e., the time to clinically significant symptom improvement, degree of symptom improvement and functional improvement, attrition, and side effect burden).
- Determine the effect of age, gender, socioeconomic status, ethnicity, and other participant features on these clinical outcomes.
- Determine the effect of the type of recruitment clinical site (e.g., primary care versus tertiary [psychiatric] care; private versus public sector) on clinical outcomes.

---

* Substantial but clinically insufficient may refer, for example, to those with a response, but who still suffer residual symptoms.
*Direct comparisons will only be made among patients who agree to randomization within and across strategic groups.

- Determine the effects of concurrent Axis I psychiatric conditions on clinical outcomes.
- Determine the effect of prior treatment history in the current episode on clinical outcomes.
- Determine the relevance of family history (as reported by participant at baseline) of response/nonresponse to specific antidepressant medication(s) on the proposed clinical outcomes.

C. Participants Eligible for Level 2A (Those Without a Satisfactory Therapeutic Response to CIT and to a Level 2 CT Switch or CT Augmentation) (See Figure 3, page 25)

C-1.   Primary Objectives: (Effectiveness)

(1) Determine which treatment switching option (switch to BUP or VEN) is associated with the greatest symptomatic improvement and remission at Level 2A.

C-2.   Secondary Objectives: (Predictors)

For each participant group defined by their Level 2 treatment, we will conduct the following secondary and exploratory analyses:

(1) Identify baseline predictors of response obtained at entry into Level 2A to each treatment option (see Section B-2 for discussion of baseline predictors).

(2) Evaluate the predictive value of early change or lack of early change in total symptom severity from baseline (e.g., baseline to week 6) for identifying those who will respond later.

(3) Identify pretreatment predictors of intolerance to the Level 2A treatment options.

D. Participants Eligible for Level 3 (Those Without a Satisfactory Therapeutic Response to CIT and to a Level 2 Medication Switch or Augmentation or to Level 2A) (See Figure 4, page 26)

D-1.   Primary Objectives: (Effectiveness)

(1) Determine the most effective treatment option(s) for individuals eligible for Level 3 treatments. The best treatment option within each substrategy group will be identified. Then, the best treatment option will be compared across substrategy groups. Direct comparisons will only be made among participants who agree to randomization within and across substrategy groups.

(2) Determine which treatment switching option (switch to NTP or MIRT) is associated with the greatest symptomatic improvement and remission at Level 3.

(3) Determine which treatment augmentation option (Li or THY) is associated with the greatest symptomatic improvement and remission at Level 3.

D-2.   Secondary Objectives: **(Acceptability and Predictors)**
For each participant group defined by their Level 2 treatment, we will conduct the following secondary and exploratory analyses:

(1) Determine which treatment options or substrategy groups (Medication Switch, Medication Augmentation) are most acceptable at Level 3.

(2) Identify baseline predictors of response obtained at entry into Level 3 to each treatment option (see Section B-2 for discussion of baseline predictors).

(3) Evaluate the predictive value of early change or lack of early change in total symptom severity from baseline (e.g., baseline to week 6) for identifying those who will respond later.

(4) Identify pretreatment predictors of intolerance to the treatment options.

E.   Participants Eligible for Level 4 (Those Without a Satisfactory Therapeutic Response to CIT and at least Three Prior Adequate Pharmacologic Treatment Trials (See Figure 5, page 27)

E-1.   Primary Objectives: **(Effectiveness)**

(1) Determine the most effective Medication Switch option for individuals eligible for Level 4.

(2) Determine which Medication Switch option (TCP or MIRT+VEN) is associated with the greatest symptomatic improvement and remission at Level 4.

E-2.   Secondary Objectives: **(Predictors)**
For each participant group defined by their Level 3 treatment, we will conduct the following secondary and exploratory analyses:

    (1) Identify baseline predictors of response obtained at entry into Level 4 to each treatment option (see Section B-2 for discussion of baseline predictors).

    (2) Evaluate the predictive value of early change or lack of early change in total symptom severity from baseline (e.g., baseline to week 6) for identifying those who will respond later.

    (3) Identify pretreatment predictors of intolerance to the treatment options.

F. Assessing the Relationship between Clinical Outcomes and Utilization and Costs by Treatment Protocol.

    F-1.   Primary Objective: **(Costs)**

    (1) Compare changes in costs of care with changes in outcomes for each treatment option (*effectiveness ratios*).

    (2) Compare costs of care across treatment options (*cost differences*).

    F-2.   Secondary Objective: **(Predictors)**

    (1) Compare costs of care by treatment options across participants by need (baseline outcome), predisposing (perceived benefits of care), and enabling factors (income, insurance status).

    (2) Compare costs of care by treatment options across provider and payer venues.

II. Background and Significance

A. Introduction

STAR*D focuses on major depressive disorder (MDD) — a common, usually recurrent, and often chronic disorder. MDD causes substantial disability and costs the U.S. over $40 billion/year (both direct and indirect costs).[75] This protocol evaluates the effectiveness of pharmacological and psychotherapeutic options in the treatment of participants with MDD who have not responded satisfactorily to an initial selective serotonin reuptake inhibitor (SSRI) — citalopram (CIT). Participants are recruited from both primary and specialty care settings.

Many treatments for depression have established efficacy in randomized controlled trials (RCTs).[61] However, the actual acceptability, clinical benefit, and side effect burden of these treatments in populations representative of every day practice is not well known. Furthermore, clinical practice guidelines suggest different levels of care for depressed patients (e.g., Crismon, et al.[28]). The strength

of the evidence for these recommendations is modest — even for those persons whose depression does not improve sufficiently even after a trial of a single antidepressant. For example, lithium (Li) augmentation of tricyclic antidepressants (TCAs) or monoamine oxidase inhibitors (MAOIs) has been evaluated,[6,102] but the efficacy of Li combined with newer agents is not as well established.[57,173] In addition, it remains unclear how to help depressed persons who respond but do not attain full remission with an antidepressant treatment. Response without full remission to antidepressant treatment is frequent[52] and associated with continuing disability and a poorer prognosis when compared to those with a full remission.*

STAR*D prospectively determines the comparative EFFECTIVENESS of different treatment options, whether used to augment the previous treatment used or as different new treatments. STAR*D aims to provide an empirical basis for practice guidelines, thereby reducing the reliance on clinical consensus. Results also provide descriptive information about selected tactics (e.g., at what dose and for how long should a treatment be conducted before declaring failure and, therefore, moving to the next level of treatment?), and describe, in the study population, the rates at which participants decline specific substrategy groups (e.g., augmentation versus switching to a new medication).

The specific treatment options will be compared using random assignment at different treatment levels. These comparisons allow us to evaluate several theoretical principles or clinical beliefs that currently guide the orchestration of treatments. For example, (1) Is it better to switch within or to switch between classes of agents? (2) Is a broader spectrum agent — one that acts directly on at least two neurotransmitter systems — more effective than a more selective agent? (3) Is switching to medications that apparently work by different mechanisms of actions (e.g., reuptake inhibition versus autoreceptor antagonism) more effective than switching to a medication with a similar mechanism of action as the prior (failed) agent? The protocol will also develop information about the effect of algorithm-driven treatment on service utilization and cost.

It is important to note that the STAR*D protocol, while extensive, will not answer every question regarding the treatment of MDD. Only one form of psychotherapy is being evaluated (though other forms will be considered once the planned protocol is completed, if NIMH continues support). To ensure sufficient homogeneity and sample size, we had to begin with a single treatment at Level 1, rather than several. Thus, the multiple roles or preferred positions in the randomization scheme for psychotherapy could not all be scientifically evaluated

---

* This protocol defines response as a clinically significant reduction in depressive symptoms (≥50% reduction in overall baseline symptom severity). Remission is the complete absence of depressive symptoms, defined for this protocol as total Hamilton Rating Scale for Depression 17-item score (HAM-D$_{17}$) of ≤7. A response without remission refers to a patient with a ≥50% decrease in baseline symptom severity at exit, but whose HAM-D$_{17}$ total score is still >7. A partial response refers to persons who do not respond but who do have clinically meaningful reductions in baseline symptom severity (e.g., ≥25% but <50% reduction in baseline symptom severity).

at this time. Similarly, several other potential first-line medication classes or types could not be evaluated.

In sum, STAR*D attempts to fill in major clinical information gaps and to evaluate the theoretical principles and clinical beliefs that currently guide pharmacotherapy. As such, it should have substantial public health and scientific significance. Findings will be obtained in representative participant groups/settings, using clinical management tools that can easily be applied in daily practice. Findings will include health care utilization and cost estimates. Results will be rapidly disseminated. These research findings should be immediately applicable to, and easily implemented in, the daily primary and specialty care practice.

B. Public Health Significance

B-1.   MDD is Prevalent: Two large, community-based epidemiologic studies in the U.S., the Epidemiologic Catchment Area Study[238] and the National Comorbidity Survey,[122] have reported a 4.9%–17.9% lifetime prevalence for MDD. In both studies, women were about twice as likely as men to suffer from MDD. These depressions occurred more frequently among youths and young adults, as well as among those with general medical conditions (GMCs).

B-2.   MDD is Typically Recurrent and Often Chronic: The course of MDD has been extensively reviewed.[32,163] For most people, MDD is a life-long episodic disorder with multiple recurrences — averaging one episode in every five-year period.[163] Both the recurrence of episodes in some and partial/nonresponse to treatment in others conspire such that approximately 20–35% of persons with MDD experience a chronic, unremitting course.[163] The DSM-IV Mood Disorders Field Trial found that 80% of those with dysthymic disorder (DD) also met criteria for a lifetime diagnosis of MDD.[118] Furthermore, data re-analysis of the ECA Study[32] found over only a one-year period that 5%–20% of persons with DD developed MDD (i.e., exhibited the so-called "double depression"). Chronicity is, therefore, a major issue, particularly among participants with double depressions, and among depressed persons with poor interepisode recovery from recurrent MDD. In fact, a study by Judd et al.[105] found that 72% of individuals with double depression and 65% with recurrent depression had chronic symptoms. Furthermore, there is a growing body of evidence that the longer the depression lasts, the more difficult it may be to treat.[105,222]

B-3.   MDD Is Very Disabling (For a review, see Depression Guideline Panel.[32,33]): The Medical Outcomes Study[239] showed that patients who meet criteria for MDD function more poorly than other primary care outpatients in three domains: limitations in physical activities, limitations in occupational or role responsibilities, and limitations in social activities due to health problems.[239] In fact, a recent World Health Organization (WHO) report[167] ranked depression as the fourth most disabling medical condition worldwide based on Disability-

Adjusted Life Years, which expresses years of life lost to premature death and years lived with a disability of specified severity and duration. They predicted that depression would be the second most disabling condition worldwide by 2020.

B-4.   The Economic Burden of Depression is High: Greenberg et al.[75] used a variety of parameters to estimate the economic burden of depression in 1990. These parameters included the cost of inpatient hospital admissions; the cost of outpatient partial care programs; and productivity lost as a result of depression-related morbidity, suicide, and other relevant parameters. They estimated that this cost in 1990 was approximately $44 billion.

Another concern is the economic burden to the health care system of depressive illness. Simon et al.[215] examined data from an HMO to identify primary care patients with depressive disorders and other patients in the same setting who did not have evidence of depressive illness. Patients with depression had almost twice the annual healthcare costs of those lacking depression. What was even more striking was that prior to the onset of their depressive illness, these patients seemed to have higher health costs, suggesting the possibility that prodromal symptoms of depression may lead to excessive use of health resources. Other interpretations are possible; for example, depression may amplify the experience of pain and lead to greater use of healthcare services, or clinicians may be more likely to pursue diagnostic evaluations in response to patients' physical and psychological distress. While the mechanism is not clearly identified, the fact remains: treatment of depressed patients entails substantial healthcare costs.

McCombs et al.[147] examined depressed patients who did not respond satisfactorily to an antidepressant treatment trial in a California Medicaid population. Treatment failure was defined as failure to refill medication prescriptions using the MediCAL database. This study found that the societal costs associated with treatment-resistant MDD were primarily due to increased hospital costs, and that these costs were $1,000 greater than for those who were not treatment failures. An obvious study limitation, however, is that treatment resistance was defined only by the pattern of antidepressant prescriptions.

B-5.   The Aim of Treatment is Symptom Remission (Not Just Symptom Reduction or Response): That symptomatic response without remission is associated with continuing functional disability,[105,106,107,156] as well as with a worse prognosis (see Rush & Trivedi[204] for a review), is well established. The indirect economic cost of failure to attain remission to antidepressant treatment is, therefore, likely to be significant.

For these reasons, acute phase treatment aims at symptom remission, not just response. The management of depressed participants who do not respond or who respond but do not attain remission with a treatment (treatment-resistant depression) is a major public health challenge, especially since ongoing, chronic

disability associated with treatment resistance accounts for a substantial portion of the overall costs involved in the treatment of depression.[148]

B-6.   Multiple Effective Treatment Options are Available for MDD: The range of treatment options available for patients with MDD has been outlined previously.[33] The AHCPR review and metaanalysis,[32,33] as well as many others (e.g., Frank et al.,[61] Jarrett & Rush,[96] Fava & Davidson[52]), reveal that about 50% of all nonpsychotic MDD outpatients initially exposed to either a time-limited psychotherapy targeted for depression or a single antidepressant medication respond to treatment, which means that the other half will continue to be symptomatic and functionally impaired after this initial treatment level.

Furthermore, of those who respond, only about 50%–65% attain remission (according to the definition of remission of Frank et al.[65]) (i.e., they no longer present residual symptoms of depression). In shorter RCTs of medications lasting 6–8 weeks, symptomatic remissions occur in about 20–30% of all those initially randomized to treatment, 20–30% respond (i.e., have a 50% reduction in total symptoms) but they do not remit (i.e., their Hamilton Rating Scale for Depression (17-item) [HAM-D$_{17}$] score is >7), 10–15% partially respond (i.e., they have a $\geq$25% and <50% reduction in baseline symptom severity), and 20–35% are nonresponders (i.e., they have <25% reduction in total symptoms) (see Fava and Davidson[52]). These numbers also appear to apply to time-limited depression-targeted psychotherapy.[191,203]

B-7.   Practice Guidelines Attempt to Organize Available Treatment Options in a Logical, Scientifically Supportable Sequence of Care: As with the rest of medicine, psychiatry now faces the challenge of recommending a sequence of treatments for depressed patients who do not respond satisfactorily to the initial treatment attempt. Psychotherapies, with efficacy established by RCTs, became available in the late 1970s (for a recent review, see Rush & Thase[203]). Furthermore, only since 1988 have antidepressant agents other than TCAs, trazodone, or MAOIs become available in the U.S.

In the face of the paucity of data directing such recommendations, it is understandable that today's evidence-based practice guidelines often do not recommend a specific sequence of treatment.[3,32,33] On the other hand, if specific sequences can be recommended based on better acceptability, efficacy, tolerability, or cost, they should reduce unnecessary practice variation, and improve outcomes. While expert consensus is a place to begin (e.g., Crismon et al.[28]), such recommendations must be prospectively evaluated scientifically to determine whether they are valid or simply current mythology.

Scientifically-based recommendations regarding treatment sequences for depression may well produce both clinical, administrative, and potentially economic benefits[72] (see also Rush et al.[197,202] for reviews of the rationale for evidence-based practice guidelines). Such guidelines must go beyond the initial

"next level" question for patients failing to respond to the first treatment because even after two levels of treatment, at best only 75% of the initial group will have responded (based on populations entering into randomized clinical trials) and less than 50% will have attained remission. Response and remission rates may be even lower in patients in routine care, given greater psychiatric and general medical comorbidities and histories of intolerance/nonresponse with prior treatments.

In addition to efficacy, a major factor in developing empirically-based treatment recommendations is the acceptability of recommended treatments to persons who suffer from MDD. For example, a slightly less effective, but far more widely acceptable treatment may be preferred over a somewhat more effective, but highly unacceptable treatment in algorithms recommending treatment sequences. No studies to date have evaluated the acceptability of different treatment options in broadly defined participant groups treated in diverse care settings.

B-8. Little Scientific Evidence Tells Us What Treatment(s) Is (Are) Recommended After the First Treatment Fails to Produce Response/Remission: For individuals who do not respond to the first treatment, a next treatment step is called for (e.g., switching to a new treatment or augmenting the first treatment with a second).

Several specific second level treatments have been subjected to research — largely open trials without comparator treatments.[58,226] In one double-blind efficacy trial comparing increased SSRI dose, Li augmentation, and the addition of desipramine to fluoxetine among more than 100 participants who had failed a prospective trial with an SSRI,[51,56,57] the overall response rate was approximately 30%, with no statistically significant differences among these three treatment options.

Furthermore, the vast majority of studies excluded participants with common general medical and psychiatric comorbidities. Thus, generalization of even these tentative findings to representative populations is difficult. The proposed study has very broad inclusion criteria that allow enrollment of both adult and elderly patients, as well as clinically depressed patients with many other psychiatric comorbidities. General medical comorbidities are also allowed, as long as clinicians consider antidepressant treatment appropriate. These design features increase the likelihood that our findings are as generalizable as possible.

B-9. How to Treat Depressed Patients For Whom Two or More Treatment Attempts Have Been Unsuccessful is Unknown: A second major group of depressed persons with a substantial public health impact are those who have not responded to multiple trials of medications and psychotherapy. These individuals may account for 10–30% of the total population with MDD. Because of the chronicity and severity of their illness, they may utilize 40–50% of treatment resources over prolonged time periods. These individuals are also more likely to commit suicide and are more likely to suffer greater morbidity or even mortality

from their GMCs.[1,91] These individuals also typically require high medication costs (because they are often treated with polypharmacy), and, in some instances, multiple psychiatric admissions. Levels 3 and 4 of STAR*D focus on the treatment on these types of participants.

B-10. Where Psychotherapy Fits in the Care of Depressed Patients is Not Clear: Millions of people with MDD receive psychotherapy alone as the initial treatment for their depression. This psychotherapy is often delivered under managed care circumstances (i.e., for a short, time-limited, duration with a cap for 5, 7, or 10 sessions – limitations that are not supported by published scientific literature[203]). Thus, we cannot be certain that depressed participants who have not responded to psychotherapy in a routine care setting have received an "adequate" course in terms of either the quantity or quality of the treatment given.

To address this issue, STAR*D permits the enrollment of participants who have historically failed to respond to psychotherapy (as long as it was not a full course — ≥16 sessions of cognitive therapy). Although less studied than pharmacotherapy, a number of reports[50,146,222] suggest that psychotherapy may play a significant role in the management of chronic or treatment-resistant depression (see Rush & Thase[203] for a review). Thus, this study compares the effectiveness of psychotherapy either alone or as an adjunct to antidepressant medication with other pharmacological treatments as "second" level treatments. In particular, one form of psychotherapy with substantial evidence of efficacy in RCTs for depression, cognitive therapy (for a review, see Rush and Thase[203]), will be tested as an augmenting treatment and as a monotherapy at Level 2.

B-11. Relapses/Recurrences: Another critical issue in the management of depressed patients is the high risk of relapses (during continuation phase of treatments; < 6 months after remission) and recurrences (during maintenance phase of treatment; > 6 months after remission),[54,117] also referred to as depressive breakthrough. Broadly conceptualized, treatment resistance can apply not only to those who do not respond to initial treatment, but also to those whose symptoms return following response or remission to acute phase treatment.[53,193]

Multiple studies show that antidepressants confer about a four-fold decrease in relapse or recurrence rates compared to placebo.[54] Nevertheless, in clinical trials between 10% to 30% of those whose acute phase treatment is associated with a substantial response, or even remission, experience depressive breakthroughs despite ongoing antidepressant treatment,[54] with relapse rates up to 57% being reported in published clinical trials.[21] If only 50% initially achieve acute remission and, with the worst case scenario, 30% of these experience depressive breakthrough within 1 to 5 years — or alternatively, 70% stay well[54,62,64] — then the proportion that get well and stay well can be as low as 35% (50% acute remission x 70% who stay well). Furthermore, long-term outcomes may be substantially worse than that found in clinical trials.[148,163]

STAR*D includes a 12-month, naturalistic follow-up of participants after response to any treatment option at any level of care. Therefore, we will establish rates of relapse and recurrence, and examine which parameters (ILLNESS, CARE, and PATIENT FACTORS) are predictive of relapse/recurrence.

B.12.   The Dissemination and Implementation of Scientifically Valid Recommendations Should Occur During the Course of This Project: Despite the high prevalence of MDD and the availability of effective treatments, underdiagnosis and undertreatment remain the norm. Only one-third to one-half of individuals with MDD are properly recognized by practitioners, and even of those who are, many do not receive adequate treatment.[89] In a study by Wells et al.[239] from data across treatment settings, only 11% of mildly depressed patients received an antidepressant and 29% of patients of high severity received an antidepressant. Of those who were treated with an antidepressant, only 41% received an adequate dose. Other studies also show that most patients prescribed an antidepressant medication are not treated at an adequate dose or for a long enough period of time. These findings indicate that there is insufficient application of research findings in the routine care of depressed patients. Finding an effective dissemination method for empirically developed evidence continues to be a great challenge.[31,46,84,143,228]

Despite tremendous resources invested in dissemination, results are often underwhelming. Simply publishing treatment guidelines or algorithms or providing them through continuing medical education (CME) programs are not especially effective in changing practice.[31] Multiple interventions and more immediate instruction or feedback on an ongoing basis have demonstrated a higher effect rate on the use of and compliance with new information and guidelines.[5,70]

Medical practice has been lagging in its use of new technologies and computers to improve patient care. A recent review shows that computer-assisted decision support systems improve both clinical performance and patient outcomes in a majority of interventions from over 68 trials reviewed.[93] To make treatment procedures more stable and predictable, health care delivery systems in the private and public sectors are making major investments in new information systems technology. These new technologies help translate the extensive, newly acquired knowledge into improved medical decisions at the time of service delivery.

In psychiatry, the last decade has witnessed a substantial growth in the number of medication alternatives with differing mechanisms of action, tolerability, and safety, as well as, in some cases, spectrum of action, for the treatment of patients with severe and persistent mental illnesses.[59,60] Results from STAR*D will provide evidence by which to inform sequential treatment algorithms — information that can then be disseminated with this new technology.

Computer technology not only provides assistance in the administrative aspects of medicine, but has also begun to assist in clinical care. A computerized treatment algorithm is a viable solution to the obstacles confronting clinicians in busy clinical settings. This approach can assist the implementing of "best practices" in various clinical environments, including primary care settings, which have significant time and resource constraints. Acceptable built-in alternatives increase adherence as this allows clinicians some freedom of clinical input.[208]

Education of both clinicians and patients is another beneficial factor in successful algorithm implementation.[78,92,192] Videos, demonstrations, and posters, along with the traditional memos from administrators and in-service training were found to assist both clinicians and patients to understand the purpose of and to develop trust in the algorithm.[16] As multiple studies show, simply providing information, either by mailing, publication, or CME, does not have a significant effect on clinician practice patterns. Stronger implementation tactics are necessary.

The implementation of the algorithm is likely to have a positive effect on patient care. Some patients may improve due to the increased attention they will receive, especially initially, through following the algorithm. Overall, patients should improve secondary to the increased degree and quality of care provided, which in the long term should decrease cost.

We will begin the process of dissemination and assist in implementation of STAR*D results as findings from the first 2 stages become available.

C. Clinical and Scientific Significance

There are several domains of essential clinical information that lack sufficient scientific evidence upon which to base specific, evidence-based, treatment recommendations for the care of depressed patients (for a review, see Rush & Prien[201]). These include answers to the following questions:

1. Which among several treatment substrategy groups produce the best clinical outcomes among depressed participants who have not responded adequately to an initial treatment? (EFFECTIVENESS)
2. Are the current pharmacological assumptions about treating depressed patients who have not responded to an initial treatment valid? Examples of such assumptions are:
   a. Switching from an SSRI to a dual action agent will be more effective than switching to another (within class) SSRI (tested as sertraline [SER] versus venlafaxine [VEN]).
   b. Switching from an SSRI to an agent affecting a different neurotransmitter system will be more effective than switching to another (within class) SSRI (tested as SER versus bupropion [BUP]).
   c. Switching to an agent (or combination of agents) with a different mechanism of action will be more effective than switching to an agent

with a similar mechanism of action (e.g., from a serotonin reuptake blocker to a norepinephrine reuptake blocker versus to an alpha-2 adrenergic auto- and hetero-receptor blocker, etc.) (tested as nortriptyline [NTP] versus mirtazapine [MIRT]) (PRINCIPLES OF TREATMENT).

3. What tactics (e.g., at what doses? over what duration of treatment?) are recommended to obtain an optimal trial of an agent before "moving on" to the next treatment? (TACTICAL ISSUES)

4. Are there any clinically useful "predictors" of response or nonresponse which should influence the choice among strategies or among specific treatment options after a depressed participant has not responded to an initial or more than one treatment attempt? (PREDICTION OF RESPONSE/REMISSION)

5. Are there clinical predictors of relapse and recurrence for depressed participants initially responding to or remitting with an SSRI? (PREDICTION OF RELAPSE/RECURRENCE)

6. What is the acceptability of various treatment strategies and substrategy groups in depressed participants who have failed an initial treatment? (TREATMENT ACCEPTABILITY)

C-1.   EFFECTIVENESS: Having determined that further treatment is unlikely to produce further benefits for an individual (in the sense that increased dose or duration of that treatment is unlikely to produce further symptom reduction), the issue becomes which among a variety of alternative treatment strategies and specific treatment options is (are) to be recommended.

A second question, embedded within the first, is whether the degree of response to the first treatment affects the treatments to be recommended for the next level (e.g., do those with some response, as opposed to those with no response to the prior treatment, preferentially benefit from psychotherapy as an augmenting treatment? Do those with no response to the first medication preferentially benefit by a switch within-class or a switch between classes of medication?).

Based largely on studies of TCAs or MAOIs or on nonmasked, nonrandomized trials, recent reviews[225,226] suggest that a switch from one medication class to another for those not responding to the first leads to response rates of 50%. However, a switch within class may be less effective. Response or remission rates for medication class switches in patients who have responded, but not remitted, to the first medication are not known.

Three uncontrolled studies[99,221,223] found that participants who did not respond to one SSRI had a 50-60% response rate to another SSRI. However, it has been suggested that markedly lower response rates to a switch within SSRIs would be observed when the failure to respond is documented prospectively, when medication doses were adjusted upwards for the initial SSRI, and when only those who did not respond to (as opposed to those who were intolerant to) the first SSRI were included. For this reason, it is not surprising that a recent survey by Fredman and colleagues (in press)[67] of 400 U.S. psychiatrists showed that the most

common next step in the management of nonresponders to an SSRI was the switch to a nonSSRI antidepressant.

Other treatment options such as augmentation strategies with either Li or thyroid hormone (see Thase & Rush[225,226]) for those who respond but do not remit, as well as apparently for those who do not respond initially (at least for Li), produces somewhere between 40–60% response rates, mostly with the older antidepressant agents. Whether these augmentation steps are equally beneficial to those who have not responded to some of the newer agents (the SSRIs) is not clear.[53] Furthermore, recent evidence[49] indicates that for those who respond but do not remit to a medication, cognitive therapy not only removes the residual symptoms, but is associated with an improved prognosis.

C-2.   PRINCIPLES OF TREATMENT: Several "theoretical" principles commonly guide antidepressant pharmacotherapy choices:

1. A switch from one agent that affects one neurotransmitter system (e.g., from a serotonin reuptake blocker) to one that affects another (e.g., selective norepinephrine reuptake blocker) will be more effective than a switch to another agent with a similar mechanism of action. This rationale has become a common basis of clinical practice, although its validity has not been studied in double-blind, controlled clinical trials.[23]

2. Agents that affect more than one neurotransmitter system (e.g., serotonin and norepinephrine) will be more effective than more selective agents (e.g., a selective norepinephrine agent) in those who have failed a more selective agent initially. For example, van Praag et al.[231] have stated that effects of both norepinephrine and serotonin are needed to optimize antidepressant effects — a hypothesis indirectly supported by the Danish University Antidepressant Group (DUAG) study,[69] but which needs empirical evaluation in representative outpatients.

3. A switch to an agent that produces changes in neuronal systems by different mechanisms of action (e.g., norepinephrine and serotonin reuptake inhibition plus post-synaptic receptor antagonism versus monoamine oxidase inhibition) than the previously failed agent will be more effective than switches to agents with similar mechanisms of action as the failed prior agent (e.g., switch from a TCA to a TCA will be less effective than switching from a TCA to an MAOI). This hypothesis has been supported by several studies,[150,151,224] but it has not been-investigated in a systematic fashion with the newer agents.[8]

In selecting among the many available effective treatments (medication, psychotherapy, electroconvulsive therapy, light therapy), we tried to balance several competing interests and priorities. Given the constraints of resources, we had to choose to answer selected questions, while choosing not to address many additional important issues. In essence, we could not study all of the available medications or psychotherapies alone or in combination. A brief synopsis of the thinking that underpins the selection of each treatment at each level follows.

Hopefully, once the STAR*D infrastructure is up and running and once the initial specific questions are answered, alternative, new treatments can be inserted and new questions can be addressed.

Level 1: Citalopram was chosen as a representative of the SSRI class. We felt that patents will end during the lifetime of STAR*D on several SSRIs which, in turn, will likely lead to managed care and other systems recommending one of these SSRIs as a first step agent. Thus, what is the "next best" step following an unsatisfactory response to an SSRI is an important question with high public health significance. Placement of citalopram in specific, and SSRIs as a group, at Level 1 should not imply that this is a preferred or endorsed practice. The selection was a research/scientific choice designed to answer one important question.

Citalopram (Level 1) and sertraline (Level 2) were selected from among the available SSRIs due to their minimal or modest effects on the CYP 450 system (i.e., minimizing drug-drug interactions), while also taking advantage of the minimal (citalopram) or modest (sertraline) proclivity to have a discontinuation syndrome — which is more likely with shorter half life SSRIs. Thus, neither agent needs, in most cases, to be tapered when switching from it to another treatment. Fluoxetine also has minimal discontinuation problems. However, its longer half-life would entail a study of a combination of fluoxetine and the next switch agent (in Level 2), which we wished to avoid for scientific reasons.

Level 2: Sertraline (see above) was selected for Level 2 to study the effectiveness of an SSRI (citalopram at Level 1) to an SSRI (sertraline at Level 2) switch. This common practice may or may not be as effective as switching "out of class" to either bupropion (which has no direct effect on the serotonin system) or to venlafaxine (which has both NE and 5HT effects). Finally, the switch to cognitive therapy (CT) is supported as a reasonable option following lack of efficacy with an antidepressant given the >50% response rate to a variant of CT called CBASP (Cognitive Behavioral Analysis Systems Psychotherapy) in patients with chronic MDD who had not responded satisfactorily to a 12-week trial of nefazodone (Keller et al., unpublished data). Thus, Level 2 switches test commonly held clinical beliefs, as well as common clinical practices. In all cases, these beliefs/practices are supported only by open, noncomparative trials.

The augmentation options for Level 2 (buspirone, bupropion, or CT) are also supported by open, uncontrolled case series (save for CT in those with residual symptoms (for reviews see Rush & Thase[203] or Rosenbaum et al.[194]). We chose CT as opposed to interpersonal or behavioral therapy because CT has more randomized trial evidence (acute phase) studies, and CT has been studied as a treatment for residual symptoms in two controlled trials.[49,180] Ideally, we would have liked to study more than one psychotherapy at Level 2 or even compared it with an SSRI at Level 1, but resource constraints required a narrower scientific focus.

Finally, we wished Level 2 to entail treatments that could be delivered in primary care settings. Thus, more complex medication regimens (e.g., lithium augmentation, combining a tricyclic with citalopram as an augmenting strategy) were excluded from Level 2.

Level 2A: Level 2A provides the same medications (as switch agents) as Level 2 for only those participants who had previously had only a single medication trial. The rationale is the same as for the Level 2 medication choices. The reasoning is that participants who enter Level 3 should have had two adequate antidepressant medication trials.

Level 3: The selection of lithium or thyroid as augmenting strategies at Level 3, which is, in most cases, will be delivered by specialists, is designed to compare two well-established augmenting agents (Li or thyroid) when used with newer antidepressants (i.e., venlafaxine, bupropion, citalopram, or sertraline). Most of the data on both lithium and thyroid augmentation rests on controlled trials with either TCAs or MAOIs (see Rosenbaum et al.[194] for a review).

The switch options in Level 3 will compare a generic tricyclic agent (TCA) (nortriptyline) with another branded agent (mirtazapine) that works by a different pharmacologic effect. Both of these agents affect both 5HT and NE systems. However, one (nortriptyline) entails a greater safety risk (cardiovascular effects) in overdose. Thus, if the newer agent is as effective as the less expensive but more medically risky agent, the newer agent would be preferred in practice. Such a finding has high public health significance.

Level 4: Level 4 participants have a relatively highly treatment-resistant depression. They will have been ill for at least 36 weeks while in the protocol (Levels 1, 2, and 3 are each 12 weeks in length). Some may also have had 12-14 weeks of CT (i.e., ill for up to 48 weeks while in the protocol). For these participants, more complex, somewhat more medically risky interactions are justified. The monoamine oxidase inhibitor tranylcypromine is one of two switch options. MAOIs appear to be effective when dual action agents such as the tricyclic antidepressants have failed.[226] The other switch option is a combination of two agents used in practice for treatment-resistant depression (namely, mirtazapine and venlafaxine). An equivalent outcome would likely recommend against the MAOI, given the dietary and other requirements surrounding its use.

In sum, the selection of the specific treatment options are an attempt to balance safety, common practice, and pharmacologic reasoning, while empirically testing choices that can be implemented in routine primary care (Levels 1 and 2) or specialty care (Levels 1-4) practice. By randomization, assessment of symptomatic and functional outcomes, as well as client satisfaction and side effect burden will inform us specifically about the particular treatment options under study. Simultaneously, post hoc analyses searching for pretreatment predictors of

response will shed light on whether particular groups of participants or types of depression might especially benefit from one or another of these options.

C-3.   TACTICAL ISSUES: STAR*D ensures that only participants who are treatment-resistant or intolerant will enter each subsequent treatment level. It is essential that participants with merely "apparent treatment resistance" to inadequate treatment do not move to the next treatment level (i.e., those without an adequate drug dose or an adequate duration of treatment at the prior level).[171] The issues of dose and duration (for both medication and psychotherapy) are termed treatment tactics.[28,200,201] Current recommendations as to dose and duration have only modest empirical support[28,155,170,184,202,218] based on post hoc data analyses.

Duration: Questions about treatment tactics will be addressed during the initial Level 1 treatment with CIT, as well as during subsequent levels. There is evidence for a biologically determined latency period that delays the onset of response to antidepressant treatment.[184] By ensuring an extensive (8-12 weeks), yet clinically practical, exposure to each treatment at each level, we will be able to determine when a treatment should be declared ineffective, as well as when its maximal benefit has been attained.

Medication Doses: By and large, clinicians tend to use inadequate drug doses.[17,18,48,101,119,123] For example, Kotin et al.,[129] Schatzberg et al.,[207] and Remick et al.[187] found that at least half the patients considered treatment-resistant and referred to specialists had received inadequate medication doses, and they responded to increased doses. In addition, many clinicians prematurely stop trials before adequate duration is met.

Studies of TCAs and MAOIs clearly support the relevance of dose.[172] Appropriately controlled studies indicate that 300mg of imipramine (or its equivalent) is statistically superior to 150mg, and that 90mg of phenelzine is superior to 45mg.[185,216,230,235] However, studies of second-generation antidepressants (e.g., SSRIs) are less numerous and less convincing as to dose-response relationships. Two studies found that a proportion of patients unresponsive to a standard dose of fluoxetine responded when a higher dose was used.[55,57] Kelsey's review[120] of several studies with VEN are consistent with a dose response curve since higher dose groups (i.e., 150–225 mg/d; 300–395 mg/d) demonstrated greater improvement than lower dose groups (75 mg/d). On the other hand, other studies have failed to show a statistically significant difference with higher doses of SSRIs.[35,182,210]

While the role of dose for second generation antidepressants is not fully studied, it is reasonable to suggest that as with first generation antidepressants, a participant should not be considered treatment-resistant unless an aggressive dose of the drug is used,[33] especially since higher doses of SSRIs are not toxic even if they only benefit a small proportion of patients.

Consequently, all medication treatment options in STAR*D will use a fixed-flexible dosing regimen with the aim of prescribing a reasonably robust dose before a drug is considered ineffective. A fixed-flexible dosing regimen has several advantages: (1) it allows the clinician to delay dose increases if the participant is initially intolerant; (2) it approximates clinical practice; (3) some participants initially intolerant accommodate to a particular dose with time thereby permitting a subsequent dose increase with an acceptable side effect burden; (4) given the range of concurrent comorbid conditions allowed into the participant study group (e.g., GMCs), some will require and tolerate lower doses, whereas others will need higher doses.

C-4.   PREDICTION OF RESPONSE/REMISSION: The degree to which the type of initial treatment, type of depression (melancholic versus atypical), demographic parameters (e.g., age, gender, or socioeconomic status), and co-existing Axis I and III disorders affects the duration of an adequate trial, as well as the likelihood and timing of response must be ascertained.

Since STAR*D strongly encourages 12 weeks of treatment with vigorous dosing at each level, it provides an opportunity to determine whether specific baseline features (Patient, Illness, Care Features) affect the likelihood and timing of clinical benefits for each treatment option, and, if sample sizes are adequate, it will determine if there are interactions between type of treatment, the above predictors, and response.

C-5.   PREDICTION OF RELAPSE/RECURRENCE: Very little is known about the long-term outcome of participants who successfully remit with an antidepressant, particularly if their remission follows non- or partial response to a prior antidepressant trial. How long are they able to remain well? STAR*D includes longitudinal, naturalistic 12-month follow-up of all participants who have a satisfactory response to any treatment option at any level. This population allows us to investigate possible clinical predictors of relapse/recurrence, as most participants who experience a relapse/recurrence during continued antidepressant treatment do so within the first 6-12 months after remission.[54,117]

How common are relapses/recurrences during continued antidepressant treatment in "real world" clinical practice? Are there clinical predictors of relapse/recurrence? Are those with greater degrees of resistance more likely to suffer more relapses/recurrences or have them earlier?

C-6.   TREATMENT ACCEPTABILITY: One of the imperatives guiding the development of STAR*D was to preserve the integral role participants typically play in negotiating treatment decisions in clinical settings. As medical practice has evolved and public access to information regarding treatment of mental illness has grown, clinicians have increasingly recognized the value of patient participation when decisions need to be made. Among patients with mood

disorders, efforts to foster patient involvement have been encouraged as a putative means to empower patients, strengthen the therapeutic alliance, optimize treatment adherence, and improve outcome.[7,27,42,63]

Given the range of options available to participants whose depression does not respond to an initial course of treatment, a participant's preference routinely wields a major influence over the selection of subsequent treatments. In clinical trials, however, scope for implementing participant preferences is usually limited. Potential participants who are reluctant to embark upon or risk randomization to a particular course of treatment typically register their preference by declining enrollment. As a result, participant pools are liable to bias with respect to such variables as past treatment history, symptom characteristics, illness beliefs and attitudes, and level of depression severity.[44,47] More generally, samples enrolled in clinical trials are unlikely to represent adequately those many patients for whom lack of influence over treatment decisions would be unacceptable. Therefore, to maximize the generalizability of this study, we believe that provisions must be made to accommodate participant preferences when they emerge.

The STAR*D design will help achieve several major goals. By incorporating acceptability, the protocol will more closely represent decision making in clinical practice than would a more standard protocol that offers randomization across all treatments as the only option. On the other hand, STAR*D preserves the power of randomization to the fullest extent feasible for each participant. We expect that this design will facilitate recruitment of a diverse, broadly representative participant population, including those individuals who might normally decline enrollment in clinical trials, and will improve participant retention by offering greater flexibility.

Finally, the STAR*D design enables us to focus upon treatment acceptability as a covariate, to address specifically whether participant acceptability for treatment substrategies, or clusters of treatments, are associated with particular clinical or demographic characteristics or have an impact on outcome. Both questions hold significant implications for the design of future studies in this area and on the delivery of optimal care.

III. Research Design and Methods

A. Overview

STAR*D aims to define the best "next level" treatment in depressed participants who have not benefited adequately from an initial (and possibly also, subsequent) treatment(s). STAR*D evaluates (1) both the acceptability <u>and</u> the comparative effectiveness of different treatment substrategies and specific treatment options for major depressive disorder (MDD) at Levels 2 and 3 the comparative effectiveness of different treatment options at Levels 2A and 4; and (2) the comparative effectiveness of different treatment options at the same levels.

Within a level, the available treatment options may be conceptualized as representing two overall treatment strategies: switching from the current treatment to another antidepressant treatment or augmenting the current treatment with an additional treatment. Within Level 2, the strategies can be conceptualized as representing two substrategies within each strategy, with each substrategy being comprised of various treatment options. These substrategies are defined by the use of medication or psychotherapy as the switching agent or the augmenting agent.

Furthermore, given that many participants will still not have attained a satisfactory response even after a second treatment trial, STAR*D will continue participants into third and fourth treatment levels — for all participants for whom the prior levels do not produce the desired clinical outcomes. The higher levels aim at those depressions that have a greater degree of resistance, having not responded satisfactorily to prior treatments.

Treatments at earlier levels are less complex, while those at later levels often are more complex (e.g., two medications used at the same time).[28] Further, where research evidence is available to suggest that specific forms of augmentation (e.g., with bupropion [BUP] or buspirone [BUS]) of selective serotonin reuptake inhibitor (SSRIs)[13,15,100,145] may be effective, or that a switch to a specific monotherapy (e.g., venlafaxine [VEN]) may be effective in treatment-resistant depression,[169] those treatment options are introduced earlier in the sequence of levels.

Please refer to Figures 1–5, which show the different treatment levels.

Figure 1: Entry and Level 1*


```
        ┌──────────────┐
        │   Obtain     │
        │   Consent    │
        └──────────────┘
               │
          Yes  │
               ▼
        ┌──────────────┐   Satisfactory        ┌──────────────┐
        │    CIT       │   Response 45%        │  Follow-up   │
        │   (4,000)    │──────────────────────▶│              │
        └──────────────┘    (1,800)            └──────────────┘
               │
               │   Unsatisfactory Response**
               │   50%
               │   (2,000)
               ▼
        ┌──────────────┐
        │              │
        │   Level 2    │
        │              │
        └──────────────┘
```
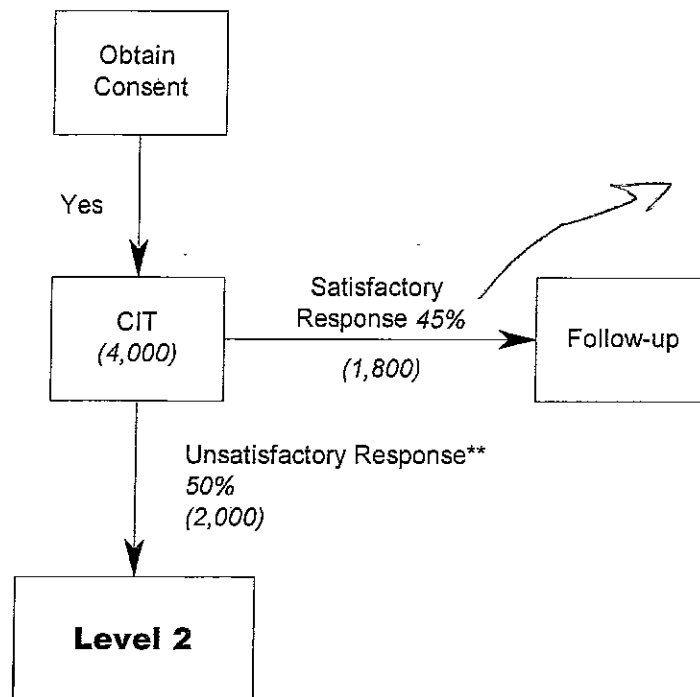
* Note: 5% (n=200) are assumed to either withdraw consent
   or to be lost to follow-up.
** Defined as nonremission.

Figure 2: Level 2

Level 1

Level 2
(for CIT nonresponders)

CIT
(4,000)

(1,800)
Satisfactory
Response
(45%)

Follow Up

Unsatisfactory Response

Obtain Consent

Strategies
Acceptable?

No → Leave Study
100 Withdraw Consent
100 "Lost"
to follow-up

Yes
50%
(2,000)

Randomize to
Options within
all
Acceptable
Strategies

SER
(287)

BUP
(287)

VEN
(287)

CT
(204)

**Switch
Options**

CIT+
BUP
(354)

CIT+
BUS
(354)

CIT+
CT
(224)

**Augmentation
Options**

**Treatment
Options**

**Treatment
Strategies**

* If strategy group is not acceptable to the patient, then patient is randomized to treatment options within remaining acceptable treatment strategies.
If all treatment strategies are rejected, then patient enters naturalistic follow-up.

Level 2

Level 2A

(for those without a satisfactory
response to CIT and to CT as
either augmenting or switching
strategies)

CT
(204)

(82)
Satisfactory
Response
40%

Follow-up

(122)

Unsatisfactory
Response

CIT + CT
(224)

(90)
Satisfactory
Response
40%

Follow-up

(134)

Obtain Consent

Med Switch
Strategy
Acceptable?

No

Leave Study
Withdraw Consent (43)
"Lost" (43)

Yes
40%
(170)

Randomize

VEN
(85)

BUP
(85)

**Treatment
Options**

**Treatment
Strategy**

Medication Switch

25

Figure 4: Level 3

Levels 2 and 2A

Level 3
(for those without a satisfactory response to Level 2 or Level 2A pharmacotherapy)

If response at Level 2, follow-up for 1 year.

Figure 5: Level 4

Level 3

Level 4
(for those without a satisfactory response to Level 3)

Satisfactory Response 30%
MIRT (174) → FU* (52)
(122)

NTP (174)
Satisfactory Response 30%
FU* (52)
(122)

SER+ Li (29) or THY (29)
Satisfactory Response 30%
F.U.* (17)
(41)

BUP+ Li (37) or THY (37)
Satisfactory Response 30%
FU* (22)
(52)

VEN+ Li (37) or THY (37)
Satisfactory Response 30%
FU* (22)
(52)

CIT+ Li (71) or THY (71)
Satisfactory Response 30%
FU* (43)
(99)

Unsatisfactory Response (488)

→ Obtain Consent →

Strategy Acceptable?
No → Leave Study
100 *Withdraw Consent*
95 *"Lost"*

Yes
41.8% (292)

Randomize

→ VEN+ MIRT (146)

→ TCP (146)

Treatment Options

* If response at Level 3, follow-up for 1 year.

STAR*D involves an initial first-level treatment with a single agent.

Rationale: We considered randomizing participants to two or more initial first-level treatments to prospectively identify participants who would need a next (i.e., Level 2) treatment. However, several first-line treatments run the risk of leading to extreme heterogeneity in the group that is eligible for Level 2. To achieve pharmacological homogeneity in the sample of depressed participants who have not responded to treatment at Level 1, we decided to expose participants to a prospective trial with a single agent. We chose an SSRI because (1) this drug group is the most widely prescribed, and (2) several SSRIs will go off patent in the next several years, making it likely that they will continue to be widely used as first line by many clinicians.

The single agent at Level 1 is citalopram.

Rationale: Citalopram (CIT) was selected to maximize feasibility. CIT currently has the lowest market share of the SSRIs with a depression indication. This should facilitate participant recruitment given a lower likelihood of prior exposure. It also has minimal discontinuation symptoms so that the tolerability of the switch to other pharmacological agents at Level 2 would not be confounded by the emergence of significant discontinuation rea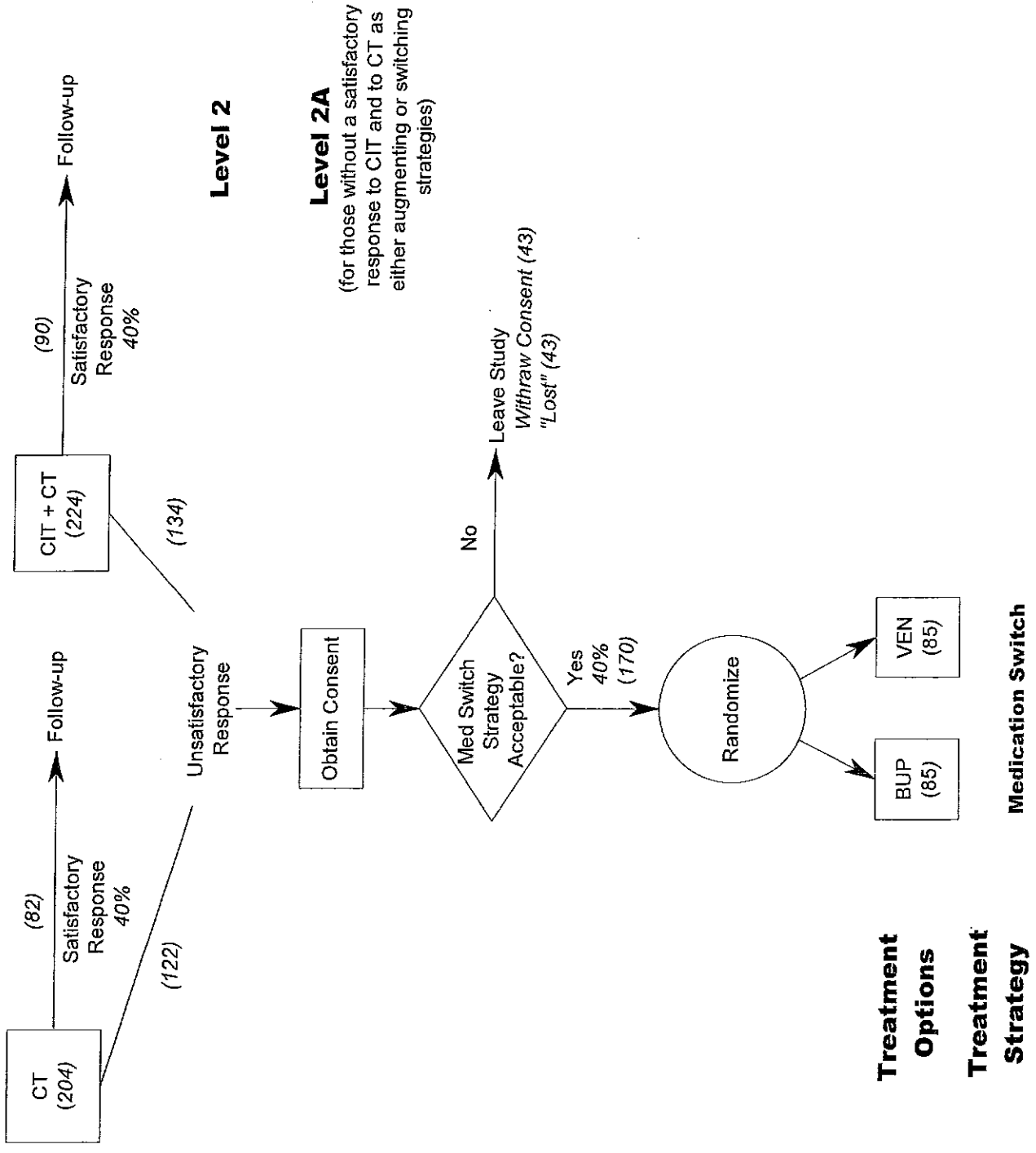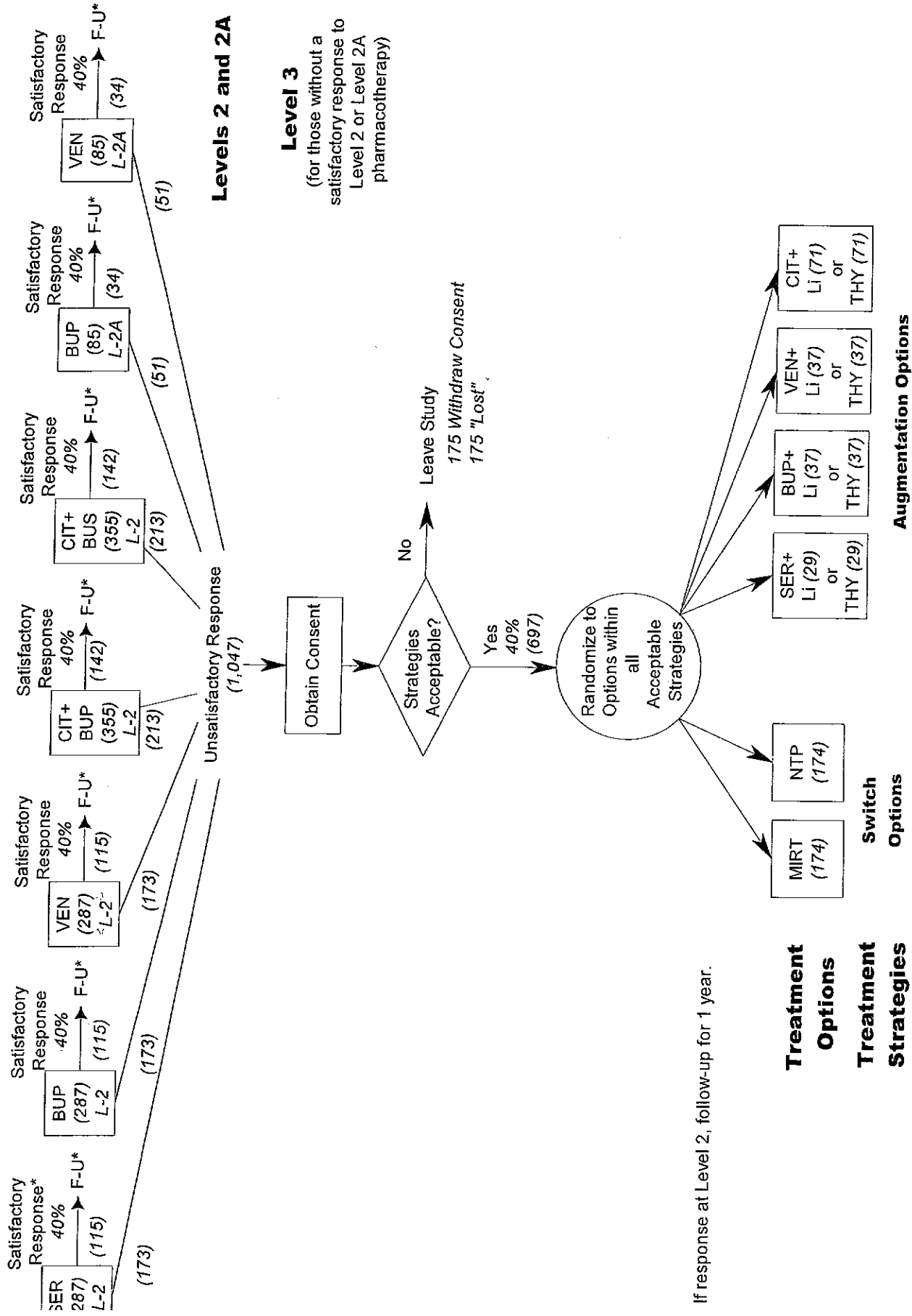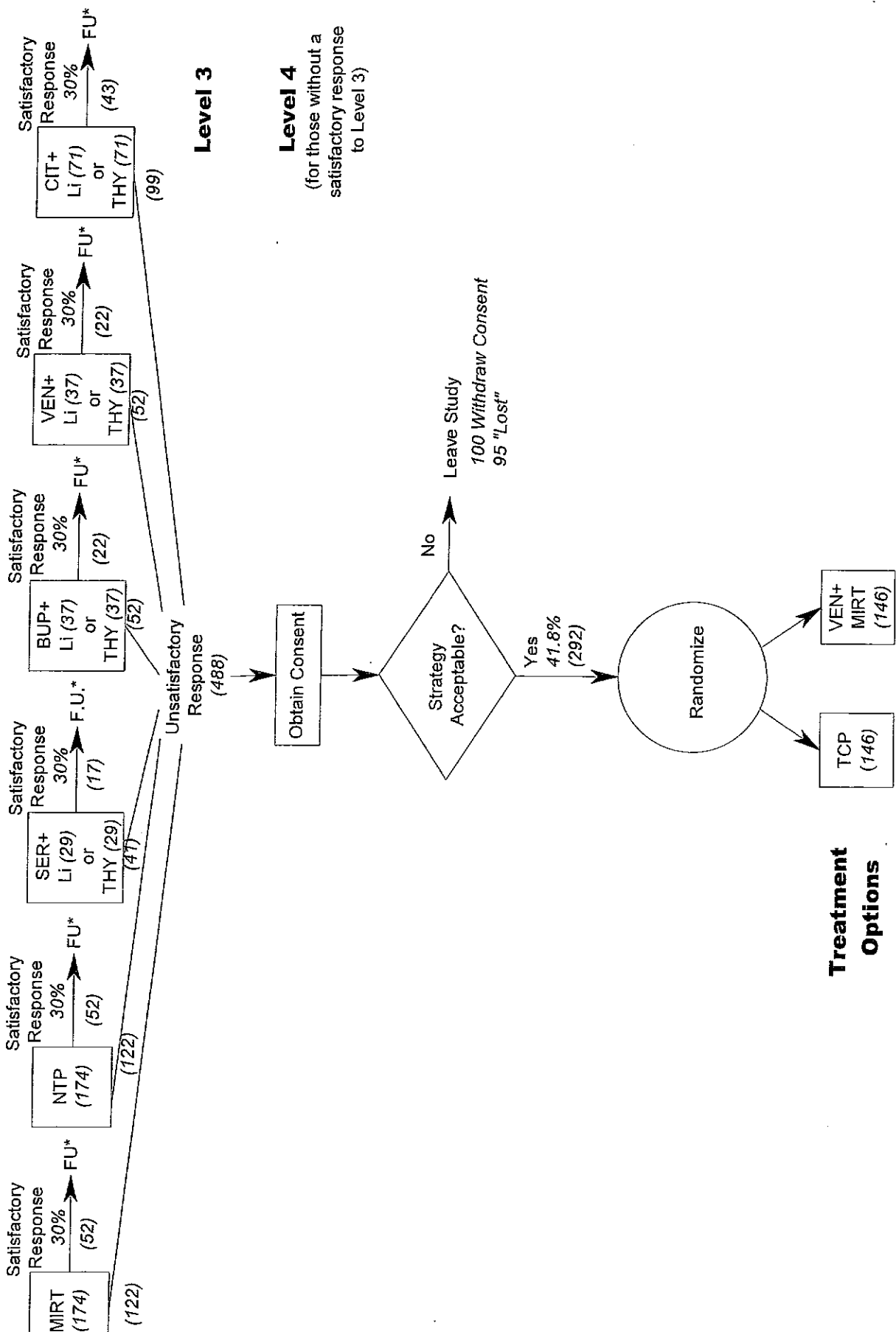ctions. CIT also has minimal risk for drug-drug interactions, thereby reducing the complications due to its use in the medically ill participants who are on concomitant medications.

The protocol strongly encourages treatment with CIT for 12 weeks using a fixed-flexible dose schedule to maximize the chances of obtaining a remission. If participants at 4 weeks have not benefited at all (defined as <20% reduction in baseline symptom severity as measured by the 16-item Quick Inventory of Depressive Symptomatology – Clinician-Rated [QIDS-C$_{16}$]), the protocol will require a dose increase to 40mg/d, side effects allowing.

Rationale: Post hoc database analyses have revealed that <20% of participants with no response (e.g., ≤20% reduction in baseline symptom severity) by week 4 will respond.

The protocol excludes participants with an established history of CIT nonresponse or intolerance in the current major depressive episode (MDE) from study entry.

Rationale: Prospectively generating all of the 2,000 participants needed to enter Level 2 ensures generalizability of the findings to both primary and specialty care clinical sites, whether delivered by publicly or privately supported clinicians.

Level 2 of STAR*D involves a novel methodological approach in which we attempt to retain a close approximation to representative clinical decision-making, while still retaining the strengths of randomized comparisons. We reasoned that participant acceptance of different strategies should have no bearing on the likelihood of response to any particular treatment strategy or option. Thus, declining a strategy or substrategy

group should not affect our ability to compare the effectiveness of the remaining (i.e., acceptable) substrategies or treatment options.

This design maximizes feasibility — often a problem in traditional randomized controlled trials (RCTs) — thereby increasing generalizability by retaining substantial flexibility for both participants and clinicians. This flexibility, we expect, will increase the likelihood of continued participation because participant acceptance is taken into account. That is, the integrity of comparisons between treatment options in various substrategies is preserved as participants will have declared acceptable all treatments within a strategy group.

There are several specific treatment options within each strategy group. This design allows participants to declare unacceptable clusters of treatments. In so doing, participants agree to randomization to all options within the strategy group(s) found acceptable (see Figure 6).

Thus, we have created "acceptability strata" (see Figure 6 for Level 2) which allow us to evaluate which overall strategies (augmentation versus switching) and particular groupings or clusters of treatment options (substrategies) are the most widely accepted. Randomization to all acceptable treatment options within clusters found acceptable allows us to define which treatments are most effective. Thus, among participants accepting randomization to all relevant substrategies, we can define the most effective treatment option within each substrategy and across different substrategies.

Finally, from an analytic perspective one might wonder whether the degree of benefit (e.g., minimal or modest) to the prior treatment might affect the relative effectiveness or acceptability of specific treatment options or particular substrategy groups. For example, could it be that those who have minimal benefits from Level 1 will differ from those with substantial (albeit still not acceptable) benefits to Level 1 in either the effectiveness or the acceptability of a Level 2 substrategy? This practical knowledge would be highly useful to inform clinicians. From an analytic perspective, the degree of benefit (from Level 1) is not likely to create a qualitative interaction between the options/substrategies at Level 2, though it may be associated with quantitative interactions.

Figure 6 shows all the mutually exclusive Acceptability Strata for the Level 2 treatments and (in parentheses) the number of participants estimated in each stratum. For example, Stratum 0 includes all participants who have agreed to medication or psychotherapy (CT) switch or augment (the "universal donor" so to speak). Those entering Stratum 1.1 have declined medication switching substrategy altogether, but are willing to accept randomization to Medication Augmentation or to CT either as a switch or augmenting substrategy. We expect few, if any, participants to accept some of these strata. For example, Stratum 1.3 would require that participants accept all augmentation options (i.e., with BUP, BUS, or CT) and to accept all medication switches, while at the same time declining a switch to CT. We estimate that few participants would select these combinations because for those with some benefit from Level 1 (CIT), we expect them to want any type of augmentation (including CT), but to decline switching to not only CT but also to the medication switches. Other combinations of treatments would, we believe, be logically far more acceptable.

Figure 6: Mutually Exclusive Level 2 Acceptability Stratum-Specific Randomizations

Stratum 0 All options acceptable	**Stratum 1.1** Med Switch not acceptable	**Stratum 1.2** Med Aug not acceptable	**Stratum 1.3** CT Switch not acceptable	**Stratum 1.4** CT Aug not acceptable
SER BUP VEN CIT+BUS CIT+BUP CT CIT+CT 8.4% (168)	 CIT+BUS CIT+BUP CT CIT+CT .7% (14)	SER BUP VEN CT CIT+CT .7% (14)	SER BUP VEN CIT+BUS CIT+BUP CIT+CT 1.1% (22)	SER BUP VEN CIT+BUS CIT+BUP CT 1.1% (22)

Stratum 2.1 Med Switch and CT Switch not acceptable	**Stratum 2.2** Med Aug and CT Aug not acceptable	**Stratum 2.3** CT Switch and CT Aug not acceptable	**Stratum 2.4** Only CT Aug or Switch acceptable
  CIT+BUS CIT+BUP CIT+CT 12% (240)	SER BUP VEN CT 12% (240)	SER BUP VEN CIT+BUS CIT+BUP 25% (500)	 CT CIT+CT 11% (220)

Stratum 3.1 Only Med Switch acceptable	**Stratum 3.2** Only Med Aug acceptable
SER BUP VEN 14% (280)	 CIT+BUS CIT+BUP 14% (280)

Total n = 2,000

* If no stratum option is admissible, participant is dropped from study.
** Two possible acceptability strata (medication augment and CT switch not acceptable and medication switch and CT augment not acceptable) have an estimated sample size of zero.

BUP = Bupropion	CT = Cognitive Therapy
BUS = Buspirone	SER = Sertraline
CIT = Citalopram	VEN = Venlafaxine

We can then estimate the number of participants assigned to each acceptable treatment option by dividing the total number of participants whom we expect to be in each acceptability stratum by the number of treatment options associated with that stratum (e.g., for Stratum 2.3, n=500 is divided by 5) to calculate the number of participants that will be randomized to each treatment option from this stratum. We then add up all participants whom we expect to enter each treatment option (i.e., summing across all strata that include the option). For comparisons between any two (or more) treatment options, the total n of relevant subjects will be the sum of those for whom both (or all) of those options are represented in their acceptability stratum.

The following sample sizes are estimated for each strategic group and for each treatment option.

Substrategy: Medication Switch

Options: VEN, SER, BUP   — 287/option
Participants come from Stratum 0 (24/option); 1.2 (2.8/option); 1.3 (3.7/option); 1.4 (3.7/option); 2.2 (60/option); 2.3 (100/option); 3.1 (93.3/option).

Substrategy: Medication Augmentation
Options: CIT+BUS   — 354
     CIT+BUP   — 354
Participants come from Stratum 0 (24/option); 1.1 (3.5/option); 1.3 (3.7/option); 1.4 (3.7/option); 2.1 (80/option); 2.3 (100/option); 3.2 (140/option).

Substrategy: Psychotherapy Augmentation or Switch
Options: CT-S   — 204
     CT-A   — 224
Participants come from Stratum 0 (24/option); 1.1 (3.5/option); 1.2 (2.8/option); 1.3 (3.7/option for CT-A only); 1.4 (3.7/option for CT-S only); 2.1 (80/option for CT-A only); 2.2 (60/option for CT-S only); 3.3 (110/option).

Obviously, these estimates are simply that, since such an experiment has not been previously conducted. However, we arrived at these estimates using three experienced practitioners ▮▮▮▮▮▮▮▮▮▮ who independently made estimates that were surprisingly close to each other. Then, via teleconferencing, the final estimates were made. The underlying assumptions of these estimates come largely by inferences from results of published RCTs.[33]

Overall, only about 50-55% of the intent-to-treat samples have a response to any medication. STAR*D participants should have no greater (and potentially lesser) response rates, given the inclusion of comorbid disorders and the enrollment of only self-declared participants (i.e., no symptomatic volunteers recruited via advertising). In

addition, a close examination of these RCTs (and our own recently completed post hoc analyses of a number of RCT databases provided by different pharmaceutical companies — including 8, 12, and 16-week acute phase trials) revealed that only two-thirds of those who respond (defined by $\geq$50% decrease in Hamilton Rating Scale for Depression (HAM-D) total score) actually attain remission (defined as exit HAM-D total score $\leq$7). Thus, complete success (i.e., remission) occurs in less than half the participants who begin any single antidepressant drug. Similar findings apply to participants treated with time-limited, depression-targeted psychotherapies, such as behavioral, cognitive, or interpersonal therapies.

We further reasoned that most participants with minimal response (or intolerance) to the Level 1 (or subsequent) treatment(s) would not likely choose to continue on the prior treatment (i.e., they would opt to switch) (e.g., Strata 2.2, 3.1). On the other hand, participants who both tolerate and realize clinically significant benefits will likely not wish to switch, but will accept augmentation with another treatment (e.g., Strata 2.1, 3.2). Further, we reasoned that some participants would decline psychotherapy, whether offered as a switch or augmentation option (Stratum 2.3).

If we have underestimated the acceptability of certain strata, then power is obviously increased (e.g., more participants are more widely accepting of more treatment substrategies). If we overestimated the acceptability of certain substrategies, then power is obviously reduced. Even if our estimates are overstated by 50%, however, nearly every substrategy still contains a sufficient sample to detect clinically meaningful differences in effectiveness. Furthermore, if some substrategies (e.g., Medication Augmentation) are rarely accepted — which we believe is very unlikely — then study resources will rightly be focused on comparing those treatments that have a good chance of acceptance in representative participant populations.

B. Treatment Levels

Before detailing the treatment options at each level, we must distinguish between clinical assessments that are conducted at each clinical visit by the clinical staff (clinician and Clinical Research Coordinator or CRC) and research outcomes assessments obtained by the Research Outcomes Assessor (ROA) or by the Interactive Voice Response (IVR) system.

For clinical visit ratings, three assessment methods to guide treatment were considered: (1) global assessments by clinicians, (2) itemized symptom self-reports by participants, or (3) itemized symptoms rated by clinicians. Our previous experiences[12] revealed only modest relationships between global ratings (by either participants or clinicians) and itemized symptom ratings by clinicians (e.g., the HAM-D$_{17}$). In addition, global ratings do not accurately reflect the complete absence of symptoms (remission), yet remission — not just improvement — guides protocol treatment in STAR*D.

We next compared itemized symptom self-reports (the Inventory of Depressive Symptomatology – Self-Report[198,199] or IDS-SR) with both global ratings by participants

and clinicians and with clinician-rated symptoms (the HAM-D_{17} and the Inventory of Depressive Symptomatology – Clinician-Rated[198,199] or IDS-C). In these unpublished, recently-completed post hoc data analyses in large samples of participants in acute phase treatment for MDD, we found that self-reported symptoms were slower to change than were clinician ratings, even when identical items were rated (e.g., IDS-C versus IDS-SR), and when the IDS-SR was compared to HAM-D. Thus, a clinician-rated, itemized symptom scale was selected to inform the clinicians who are making decisions about medication dosing and about suitability of participants to enter the next level of treatment or to enter follow-up.,

Three clinician ratings were considered: the HAM-D_{17}, , the IDS-C_{30}, and the 16-item Quick IDS (QIDS-C_{16}). The QIDS-C_{16} was selected because (1) it contains all 9 criterion symptoms of MDD (including suicidality, which is not rated by the HAM-D_6); (2) it is shorter than the HAM-D_{17} or IDS-C_{30}; and (3) it is not identical to our primary Research outcomes assessment of symptoms — the HAM-D_{17}. In the following discussion of levels of treatment, we refer, therefore, to the QIDS-C_{16} as the measure used by clinicians to gauge participant response/remission at each clinical visit. The research evaluation of effectiveness will rest on the HAM-D_{17} obtained, not by the clinician or CRC, but by telephone interview with the ROAs.

Note: We have conducted data analyses (as yet unpublished) that provide for converting various HAM-D_{17} thresholds (e.g., remission, defined as ≤ 7) to QIDS-C_{16} thresholds. These analyses provide a basis for specifying interim QIDS-C_{16} targets at various critical decision points (e.g., weeks 4, 6, and 9) in the implementation of each treatment as specified in the Clinical Procedures Manual.

Overview: Eligibility for each next level or for follow-up depends on the symptom status of the participant at the end (exit) of the level. The STAR*D protocol strongly encourages going to the next treatment level (rather than to follow-up) for both nonresponders, as identified at the clinic visit with the QIDS-C_{16} (nonresponse is <50% reduction), and for responders who do not attain full remission (defined clinically by QIDS-$C_{16} \leq 5$).

Rationale: Remission is the target of treatment. A full, robust trial (12 weeks with adequate dosing) will have been conducted at each prior level. Both nonresponders and responders who have not attained remission will have sufficient symptoms to warrant another treatment step.

However, participant and clinician judgment must enter into this decision. So a participant may opt to not take the next step, even after having been provided information about the desirability of remission. Participants who respond and who remit prior to the last visit and have sustained the remission at week 12 may enter the 12-month, naturalistic follow-up. A participant who only initially showed a response or remission at week 12 may have another treatment visit at that level after another 2–3 weeks to determine if the response or remission status is sustained.

<u>B-1.</u>   <u>Level 1</u>: We will enroll 4,000 outpatients with MDD into Level 1 over a 24-month enrollment period (4,000 participants/24 months = 167 participants/month). All participants will receive CIT 20–60 mg/day for 12 weeks at Level 1 to gauge the full benefit of the treatment, except for those with clear intolerance, or those with no significant reduction in baseline symptom severity (<15% by week 6 or <25% by week 9). If <20% reduction in baseline symptom severity is found at week 4, the initial dose (20mg/d) will be raised (assuming tolerable side effects) to 40mg/d. Similarly, at week 9, dose will be raised again (given side effect tolerability) to 60mg/d. Participants advance to Level 2 if no change has been found at 2 weeks on this maximal dose, or if remission has not occurred after ≥6 weeks on the maximal tolerated dose.

At the end of Level 1 acute treatment (12 weeks), participants with remission (or response without remission who cannot be encouraged to proceed to the next level) enter follow-up and continue treatment with the same dose of CIT. Research outcomes assessments are obtained every 3 months during follow-up. Participants are seen by their clinicians every 2–3 months in accordance with accepted guidelines.[32,33] Follow-up lasts for 12 months from the time of entry into follow-up from any level of treatment.

<u>B-2.</u>   <u>Level 2</u>: If the response to CIT (Level 1) is not satisfactory (or if participants have unacceptable side effects), then Level 2 treatments are offered. We expect approximately 2,000 participants to be eligible for Level 2. Level 2 includes seven individual treatment options grouped together within six substrategies (Figure 2) and acceptability strata (Figure 6). Medications included as switch options are SER, VEN, or BUP. Medication augmentation treatment options include BUP or BUS given in addition to CIT. Psychotherapy (cognitive therapy) is available as either a switch or augmenting treatment.

Participants consenting to Level 2 treatments will be block randomized, stratified by clinical site and preference stratum. Participants are automatically randomized to Level 2 treatment switch options at entry into Level 2 if they are intolerant to CIT.

<u>Intolerance</u> is declared if a participant discontinues CIT during the first 4 weeks for any reason or encounters intolerable side effects that require the medication to be stopped at a point subsequent to the 4th week of treatment (independent of the symptomatic status).

<u>Nonresponse</u> is <50% reduction in the QIDS-C_{16} at exit, as long as the exit is not due to intolerance.

<u>Response without remission</u> is ≥50% reduction in baseline QIDS-C_{16} score, and a QIDS-C_{16} score that is >5 at exit from CIT, as long as the exit is not due to intolerance.

Figure 7: Expected Participant Flow through Levels of Treatment

Levels	Results	Disposition
1 (4,000 enter)	Satis. Resp. (45%) (1,800)	Follow-up
	"Lost" (2.5%) (100)*	NA
	Withdraw Consent (2.5%) (100)	TAU
	Agree to L-2 (50%) (2,000)	L-2
2 (2,000 enter)	Satis. Resp. (40%) (800)	Follow-up
	"Lost" (10%) (200)	NA
	Withdraw Consent (10%) (200)	TAU
	Agree to L-2A (8.5%) (170)	L-2A
	Agree to L-3 (31.5%) (630)	L-3
2A (170 enter)	Satis. Resp. (40%) (68)	Follow-up
	"Lost" (10%) (17)	NA
	Withdraw Consent (10%) (17)	TAU
	Agree to L-3 (40%) (68)	L-3
3 (698 enter) [68 from L-2A] PLUS [630 from L-2]	Satis. Resp. (30%) (209)	Follow-up
	"Lost" (13.6%) (95)	NA
	Withdraw Consent (14.3%) (100)	TAU
	Agree to L-4 (41.8%) (292)	L-4
4 (292 enter)	Satis. Resp. (25%) (73)	Follow-up
	Nonresponse (75%) (219)	Study Exit

* Lost = Leaves treatment entirely. Cannot be found.
TAU = Treatment as Usual

Even though approximate, these numbers show our expectation that a very substantial number of participants will not benefit from Level 2 or even Level 3. We have generously estimated response rates at each level based on the RCT literature that reveals 30–40% remission rates in intent-to-treat samples (see Thase & Rush[226]). However, participants in this study will have more concurrent general medical conditions (GMCs) and other psychiatric comorbidities than participants typically entered into efficacy RCTs. Therefore, they are likely to benefit even less dramatically than persons in efficacy RCTs (see Keitner et al.,[116] Depression Guideline Panel[32,33]). Recall that the STAR*D protocol defines remission, not just response, as the aim of treatment. Thus, a "satisfactory therapeutic response" in most cases should be symptomatic remission.

If remission is not attained at Level 1 (or if participants have unacceptable side effects), then the Level 2 treatments are offered. The acceptability of each treatment strategy/substrategy requires that all treatment options within that strategy are acceptable. If a participant excludes the possibility of being assigned to a treatment option within a

treatment strategy or substrategy group, he/she will be randomized only to the remaining treatment substrategies where all options (within the substrategy group) are viewed as acceptable treatments by the participant. Therefore, participants will be randomly assigned to all options within all the treatment clusters that they deem acceptable. (Note: Participants cannot select among the medication options within either the switching or augmenting substrategy groups.)

For example, participants who reject both Medication and Psychotherapy Switching, but who do accept Medication or Psychotherapy Augmentation — will be randomized to all 3 treatment options within this strategy (i.e., CIT plus BUS; CIT plus BUP; CIT plus CT). Therefore, all participants who are eligible to enter Level 2 must accept being randomized to at least one of the six treatment substrategy groups (Figure 6).

B-3.   Level 3: Participants without remission move to Level 3. Again, assignment at Level 3 will be prospectively stratified based on clinical site and preference stratum.

Level 3 treatment groups include: 1) Medication Switching (either mirtazapine [MIRT] or nortriptyline [NTP]); 2) Medication Augmentation (either lithium [Li] or thyroid hormone [THY]). At the conclusion of Level 3, participants are again classified based on the degree of benefit from the Level 3 treatment option.

B-4.   Level 4: Participants not in remission from Level 3 are eligible to move to Level 4. Level 4 treatments are implemented only after at least a 7 day washout period for medications from Level 3, since a monoamine oxidase inhibitor (MAOI) is used in Level 4. Assignment at Level 4 will be prospectively stratified by clinical site. Treatments include only medication switches to either tranylcypromine (TCP), an MAOI, or to the combination of VEN and MIRT.

B-5.   Follow-up: All participants who remit after any level of treatment will be followed with full research outcomes assessments every 3 months and IVR assessments monthly for a total of 12 months. Clinic visits are based on clinician judgment but are expected to be held at least every 3 months. All participants will be encouraged to continue on the same type and dose of medication that was successful in the acute phase. Participants in cognitive therapy are required to continue the treatment with a minimum number of booster sessions when they enter follow-up (they must have either a response or a remission to enter follow-up). Additional sessions will be required in the case of response without remission, when they enter follow-up.
Research outcomes assessments by ROAs will be identical to those conducted at entry into and exit from each level of acute phase treatment (i.e., the HAM-D$_{17}$ and IDS-C$_{30}$). Furthermore, IVR will be used to obtain identical ratings of function, side effect burden, participant satisfaction, and health service utilization information as those obtained at entry into and exit from each level in the acute phase (see below).

C. Participant Selection

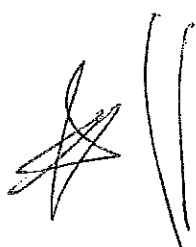This section describes the characteristics of eligible participants.

C-1.   Recruitment: To achieve the goal of recruiting a broadly representative group of participants with MDD, clinical sites will be selected from groups providing primary and specialty care in either the public or private sectors. To further ensure that the study sample is representative of the "real world," clinical sites will include practice sites that do not typically engage in traditional RCTs. Further, advertising for the purpose of recruiting participants will not be permitted, since it tends to enroll a less representative spectrum of participants.[11,211] At many sites with high participant volume, more participants will be eligible than can be entered. In those sites, we will ensure that clinicians <u>randomly</u> approach eligible participants (as opposed to selecting participants whom the clinicians believe are most likely to benefit from the protocol) by following the procedures outlined in the National Coordinating Center (NCC) and the Regional Center (RC) Manuals.

C-2.   Inclusion/Exclusion Criteria: In general, the inclusion/exclusion criteria are broad so as to acquire a sample representative of persons with MDD who would receive medication or psychotherapy in everyday practice. However, persons with medical contraindications that preclude randomization to any treatment in Levels 2-4 are excluded. In addition, participants with schizophrenia, schizoaffective disorder, bipolar disorder, anorexia nervosa, bulimia, or obsessive-compulsive disorder (OCD) as a primary diagnosis are excluded because they have a primary psychiatric condition that requires a different initial treatment.[32] Participants with currently active and clinically significant substance abuse are eligible (so long as inpatient care is not required clinically at study entry), though participation in a substance abuse program will be encouraged by their clinician. Participants with active substance dependence who require detoxification are not eligible for reasons of medical safety.

Inclusion criteria:

1) Outpatients with nonpsychotic MDD.

Rationale: Research has demonstrated high 12-month prevalence rates for MDD in community samples. MDD has significant disability, morbidity, and mortality (see Section IV). Including minor and other forms of depressive disorder not otherwise specified (DNOS) risks the enrollment of a very heterogeneous population with a course of illness yet-to-be well characterized, which would confound results with higher-than-expected rates of spontaneous recovery (nearly half of those with DNOS remit spontaneously).

2) A score of ≥ 14 on the HAM-D$_{17}$.

Rationale: Medication exceeds the effect of placebo in primary care participants with HAM-D$_{17}$ ≥ 12.[178] We added 2 HAM-D$_{17}$ points to take into account the possibility of measurement error.

3) Outpatients for whom antidepressant treatment is deemed appropriate by the treating clinician.

4) Age range: 18–75.

   Rationale: SSRIs have shown to be efficacious in adolescents with MDD and the public health significance of disorders in adolescents is high. However, NIMH has already awarded a multi-center contract to study the treatment of depressed adolescents. In addition, the psychotherapeutic approach used in this protocol (cognitive therapy) requires both significant modifications for this population and specific expertise in providing this type of therapy to participants in this age range, greatly limiting the feasibility of recruiting representative clinical sites which include practice sites not typically engaged in traditional RCTs. We anticipate that, for elderly persons >75 years of age, concomitant GMCs make a number of the treatments in the protocol medically risky or even medically contraindicated.

5) Participants with suicidal ideation are eligible, as long as outpatient treatment is deemed safe by the clinician (i.e., inpatient care is not called for clinically).

   Rationale: There is no reason to think there is a better available treatment for depressed participants with suicidal ideation than the ones at each level. Indeed, recent evidence suggests that most such participants receive inadequate antidepressant treatment in community settings.[174,175] Thus, protocol participation is not a greater risk for these participants than usual care. Participants who become so suicidal that inpatient care is needed will be removed from the study.

6) Participants who have most GMCs are eligible. Participants whose GMCs could conceivably be physiologically causing their depressive symptoms will receive treatment as usual for their GMCs as well as protocol Level 1 treatment for their MDD. We anticipate that during Level 1 most medical GMCs will be treated so that if depressive symptoms persist after treatment for the GMC and after CIT for the depression, participants with these conditions are eligible for randomization into Level 2.

Exclusion criteria:

(1) Participants must not have an established, well-documented history of nonresponse or clear intolerability in the current major depressive episode to one or more treatments required by the protocol, delivered at an adequate dose (e.g., $\geq$40mg/d of citalopram for at least 6 weeks or $\geq$ 16 sessions of CT).

   Rationale: By prospectively studying participants who are not already extremely treatment resistant and ascertaining the level of resistance, we (a) are sure of the finding, and (b) will be developing treatment steps applicable to a large and widely representative group of patients.

(2) Participants with a lifetime history of bipolar disorder (BPD I, II, and NOS), schizophrenia, schizoaffective disorder, or MDD with psychotic features.

(3) Participants who currently suffer from a primary diagnosis of anorexia nervosa, bulimia nervosa, or obsessive compulsive disorder (OCD).

(4) Participants with severe, unstable concurrent psychiatric conditions likely to require hospitalization within six months from study entry (e.g., participants with severe alcohol dependence who have a history of recent admissions aimed at detoxification).

(5) Participants with substance dependence disorders who require inpatient detoxification. Participants with active substance abuse or dependence disorders who enter the study will receive whatever care is routine at their clinical site (e.g., substance abuse counseling) for these conditions.

(6) Participants with certain concurrent psychiatric or medical conditions that are relative or absolute contraindications to the use of more than one treatment option within the protocol so that randomization to any of the strategies or substrategies within each level (Levels 2, 3, and 4) is not possible. Participants with certain concurrent psychiatric or medical conditions that are relative or absolute contraindications to the use of one or more of the treatment options within the protocol and with the possibility of randomization to at least one of the strategies or substrategies within each level (Levels 2, 3, and 4) may enter the study, as long as the contraindication is noted and the strategy/substrategy involving the contraindicated treatment option is dropped.

(7) Participants taking any concomitant nonpsychotropic medications (save for anxiolytics and sedative hypnotics) may enter the study as long as their clinician determines that antidepressant treatments in the protocol are appropriate and safe. When there is a known association between the concomitant medication and depression, as suggested by the AHCPR guidelines, we will encourage clinicians to substitute, whenever possible, the concomitant medication with another that is not associated with depression before study entry or, when the latter is not feasible, during Level 1.

(8) Participants already receiving a targeted psychotherapy aimed at their depression may not enter the study. Those who have not responded to such psychotherapy and subsequently terminated it prior to study enrollment or those who are receiving counseling or therapy for other problems (e.g., marital counseling to address marital discord; psychodynamic treatment of character issues) may enter the study.

(9) Participants who are pregnant or who will be trying to become pregnant within the subsequent 6-9 months.

C-3.   Ethnic and Gender Issues: Every effort will be made to recruit a broad spectrum of participants representing all racial groups and both genders. The study protocol will be

developed according to the principles of the Declaration of Helsinki. All risks, benefits, and adverse events associated with each treatment within the randomized treatments will be explained to study participants, who must provide written informed consent prior to study participation.

D. Randomization (Stratification/Blocking)

All participants who are eligible, who consent to participate in the trial, and who do not remit to Level 1 treatment (CIT) undergo randomizations at Levels 2, 2A (when indicated in participants who do not remit at Level 2), Level 3 (for those who do not remit at Level 2 or 2A), and Level 4 (for those who do not remit at Level 3). As mentioned earlier, the integrity of randomized comparisons is preserved by analyzing, for comparative purposes only, treatment options within groups of treatments found acceptable by participants (see Acceptability Strata in Figure 6).

Participants enrolling in Levels 2, 2A, 3 and 4 will be block randomized, stratified by clinical site. In addition, those enrolling in Levels 2 and 3 will also be stratified by the pattern of treatment substrategies that are acceptable.

D-1. Randomization Process: Participants eligible and consenting to participate in the trial will be assigned treatment at random using an IVR system. The CRC or clinical staff at each clinical site will be able to call a protected area of the IVR system accessible only to study personnel.

Relevant information will be collected via the IVR system including the names of the clinical site and the study personnel at the clinical site performing the randomization, the level of randomization, type of treatment failure, and the acceptability stratum chosen by the participant. The name of the participant will not be entered into the database. Once all the data have been entered, the treatment option will be assigned. Data entered into the IVR system a fax will be sent to the NCC, the EDC, and to the study personnel that randomized the participant notifying them of the participant's entry into the study, the stratum to which the participant was assigned, and the treatment option assigned (this last item is not sent to the NCC).

If the data entered do not meet the on-line edit checks (e.g., a participant failing to meet all the inclusion and exclusion criteria), no random treatment assignment will be assigned. The clinical site will be notified immediately of the inconsistency and asked to correct the problem. Data, once entered, are transferred and stored in a database at the Data Center (EDC).

D-2. Random Treatment Assignment: Uniform random number generation is used to randomize participants to the treatment options within each acceptability and clinical site stratum.

Randomization of participants will be blocked based on the number of treatment options available within an acceptability stratum. The advantage of this is that after a specific

number of participants have entered the trial, a balanced design is assured. The disadvantage is that for a block size of n, once n-1 participants have been randomized, the treatment of the nth participant is then known.

E. Clinic Visit Schedule

Participants will be seen at weeks 0, 2, 4, 6, 9, and 12 at all levels that entail medication visits. At subsequent levels, the exit of the prior level is considered week 0 at the next level, if the visit for the new level occurs within 3 days of the exit from the prior level. The visit schedule is flexible in the following ways. For example, the week 2 visit may be held within $\pm$ 3 days of week 2. Extra visits, if clinically needed, may also be held. If a participant has a recent onset of either response or remission at week 12, an additional visit should be held at that level of treatment to determine if the status at week 12 is sustained.

Ideally, psychotherapy visits are to be twice a week for weeks 1-4, then once a week for 8 weeks. If nonresponse is present at visit 16 (end of week 12), participants proceed to the next level. If response without remission is present, the therapist may continue for an additional 4 weeks with weekly sessions, followed by 4 every other week sessions and 6 monthly sessions, for up to a total of 30 sessions. Participants who remit before week 12 (16 sessions) and who have at least 3 consecutive weeks of remission may go to follow-up without completing all 16 visits. However, the therapist may schedule up to 4 additional booster sessions over the next 12 months.

F. Dosing Schedule (see Appendix A)

Medications are dosed with the recommended ranges primarily based on package insert. Aggressive dosing, with a balance in terms of side effects, is encouraged.

G. Protocol Adherence at the Treatment Clinical Sites

The implementation of the treatments to be evaluated must be well conducted and representative of good practice, yet they must be tailored to individual participants who are medically, ethnically, and demographically diverse. This tension between generalizability and protocol adherence results in a need for a spectrum of clinically acceptable variations (i.e., variations that do not profoundly affect outcomes or that affect outcomes equally across all treatment options to be compared).

Adjunctive Treatments: Adjunctive treatments are treatments used to manage transient associated symptoms (e.g., insomnia) or transient or longer-term medication side effects (e.g., sexual dysfunction). Adjunctive treatments cannot be ones aimed at treating depression or cannot be treatments that are used clinically as adjunctive treatments of antidepressants in treatment resistant depression (e.g., buspirone, bupropion). The overarching principles guiding adjunctive treatments try to balance the need to minimize the confounding effects of treatments that may alter the effect sizes of treatments under study with the goals of increasing the feasibility of conducting a randomized treatment

protocol and of increasing treatment adherence among participants. By allowing adjunctive treatments that are widely used in practice and that are acceptable to participants, STAR*D enhances both generalizability and feasibility.

Concurrent Psychiatric Medications: At study entry and during the study, antipsychotic (e.g., typical and atypical antipsychotic drugs) and mood stabilizing agents (e.g., lithium, carbamazepine, valproate, lamotrigine, gabapentin, or other anticonvulsants) are excluded. If participants have been taking inadequate doses of antidepressants, these medications will be tapered off and stopped, after consent but before beginning Level 1 treatment. Baseline assessment for entry into Level 1 begins when (within 72 hours following) citalopram is started for treatment protocol participants who enter the treatment sequence. During protocol treatment, benzodiazepines and imidazopyridine hypnotics are allowed.

Concurrent Nonpsychotropic Medications: In general, participants can enter Level 1 receiving any medication(s) for concurrent GMCs as long as there is no contraindication to use with any medications in the STAR*D algorithm.

Detecting Protocol Deviations and Violations: Protocol <u>violations</u> are deviations from protocol that destroy the internal validity of the study (e.g., use of antidepressants not included in the protocol). All other deviations are <u>deviations</u>, not violations. The Clinical Procedures Manual specifies acceptable clinical variations in the protocol medication treatments. The Psychotherapy Manual specifies similarly acceptable clinical variation in the use of cognitive therapy.

The two main methods for detecting both deviations and violations are (a) the Clinical Record Form (CRF) completed at each visit and (b) the weekly, biweekly, or monthly supervisory teleconferences conducted with the clinicians.

The CRF is the mainstay for detecting protocol deviations and violations. At each visit, clinicians use the CRF to record: current medications, other current treatments for depression, and other current treatments for GMCs; clinical status defined by depressive symptom severity (QIDS-C_{16} total scores); clinician (Clinical Global Impression – Improvement [CGI-I][79]); and participant global ratings of side effect frequency, intensity, and burden (FISER/GRSEB). Participants also complete an 8-item self-report (the Patient Rated Inventory of Side Effects [PRISE]) at each visit. Clinicians note severity of any associated symptoms. Clinicians record the medications and any other treatments prescribed at the conclusion of the visit.

The CRF contains all of the information needed to detect protocol deviations and violations (i.e., failing to raise the dose appropriately in the face of persistent symptoms, or prematurely moving or delaying the move to the "next step"), since it records dose, symptom severity, and side effect information. The CRF also records each participant's degree of adherence to dose and medication type. If participant adherence is low, the CRC or clinic staff provides counseling education (see the Patient Education Manual).

The CRC reviews the CRF during each clinic visit. Violations are brought to the attention of the clinician and the Regional Center (RC) Director, if needed. The RC Director contacts the clinician if remedial action is needed.

On the teleconferences, the CRC from each RC and the NCC Co-Director review the protocol rules and ask about deviations. Case-based discussions follow. Continuing medical education (CME) credits are provided for these teleconferences.

Protocol violations impair the integrity of the study. Study monitoring by the EDC will detect protocol violations, which will be communicated to the NCC Co-Director. Each violation will be brought to the attention of the RC Director and the clinician. For each protocol violation, a form to document the nature of the violation will be filled out by the clinician.

The RC Director, initially, with the Clinical Manager at the NCC, reviews the violation to determine if it is serious enough to drop the participant (e.g., a participant with history of nonresponse to adequate treatment with CIT or SER during the current episode was entered into the study) or is "minor" (e.g., participant prescribed an atypical antipsychotic for 2 days to help with anxiety).

When a clinician or clinical site accumulates over 5% serious protocol violations, the NCC Director, with the STAR*D Executive Committee (EC), determines whether or not the clinical site should continue. If remedial measures fail to rectify the problems, the clinical site is dropped.

Preventing Protocol Violations by Participants: The importance of adherence to protocol procedures is emphasized to participants with an educational package that includes information about the nature of MDD, simple theories of mechanisms of action of antidepressants and of cognitive therapy — with an emphasis on adherence to prescribed antidepressant medications, or homework for those in cognitive therapy, and clarification of reasonable expectations about the timing of therapeutic effect. Further education will focus on early and longer-term medication side effects, noting that most side effects abate within a reasonable period of time. An educational package has been developed to cover the above issues. In addition, a monthly newsletter will be mailed or emailed to each study participant to maintain interest and motivation in STAR*D. We will also provide the same newsletter information on the STAR*D website.

Preventing Protocol Violations by Clinicians: RC Clinicians will be trained at the STAR*D START-UP meetings in Dallas, at which RC Directors, Associate Directors, and CRCs will be trained in the protocol. They, in turn, will train the clinicians at their Clinical Sites. Ongoing teleconferences (weekly->biweekly->monthly) will be held over the first year of participant entry with clinicians, CRCs, and the NCC Co-Director and Clinical Manager to check on/supervise protocol implementation. Each clinician is also provided with a Clinical Procedures Manual detailing all clinical and relevant research

procedures. Additionally, a checklist of inclusion and exclusion criteria is completed by the CRC for each entering participant to ensure consistency with the protocol.

H. Baseline Assessments

The following discussion refers to information found in Tables 1 and 2.

Table 1. Schedule of Screen and Baseline Assessments**

Screening Assessments						
Domain	**Measure**	**Time**	**How**	**When**	**Who**	**Where**
Intake						
Consent	Consent	20 min	Interview	At Intake	CRC	Clinic
Character-istics	Clinical/ Demographic Features/ Social Support	4 min	Interview	At Intake	CRC	Clinic
Eligibility	Inclusion/Exclusion	5 min	Interview	At Intake	CRC	Clinic
Diagnosis	PDSQ	20-30 min	Self-report	At Intake	Participant	Clinic
Symptoms	HAM-D_{17}/QIDS-C_{16}	15 min	Interview	At Intake	CRC	Clinic
GMCs	CIRS	5 min	Interview	At Intake	CRC	Clinic
Baseline Research Outcome Assessments						
Symptoms	HAM-D_{17} /IDS-C_{30}	20-25 min	Telephone Call	Within 72 hrs of Intake	ROA	Home/Work
	$QIDS_{16}$	6 min	Telephone	Within 72 hrs of Intake	IVR	Home/Work
Function	SF-12	4 min	Telephone	Within 72 hrs of Intake	IVR	Home/Work
	WSAS	2 min	Telephone	Within 72 hrs of Intake	IVR	Home/Work
	WPAI	2.5 min	Telephone	Within 72 hrs of Intake	IVR	Home/Work
Quality of Life	Q-LES-Q	6 min	Telephone	Within 72 hrs of Intake	IVR	Home/Work
Side Effects	N/A	N/A	N/A	N/A	N/A	N/A
Participant Satisfaction	N/A	N/A	N/A	N/A	N/A	N/A
Utilization & Cost	UAC PQ-15	5 min	Telephone	Within 72 hrs of Intake	IVR	Home/Work
Income	IPAQ	3 min	Telephone Call	Within 72 hrs of Intake	ROA	Home/Work

**Screening and baseline research outcomes assessments are obtained on all participants (n=4,000).

GMCs = General Medical Conditions
IVR = Interactive Voice Response
ROA = Research Outcomes Assessor

CIRS = Cumulative Illness Rating Scale
HAM-D_{17} = Hamilton Rating Scale for Depression (17-item)
IDS-C_{30} = Inventory of Depressive Symptomatology – Clinician-Rated (30-item)
IPAQ = Income and Public Assistance Questionnaire (5-item)
QIDS-C_{16} = Quick Inventory of Depressive Symptomatology – Clinician-Rated (16-item)
$QIDS_{16}$ = Quick Inventory of Depressive Symptomatology (16-item)
Q-LES-Q = Quality of Life Enjoyment and Satisfaction Questionnaire (16-item)
PDSQ = Psychiatric Diagnostic Screening Questionnaire
SF-12 = Short-Form Health Survey (12-item)
UAC PQ-15 = Utilization and Cost Patient Questionnaire (15-item)
WPAI = Work Productivity & Activity Impairment Questionnaire (6-item)
WSAS = Work & Social Adjustment Scale (5-item)

Table 2. Research Outcomes Assessment Battery*

Domain	Measure	Time	How	Who	Where
Symptoms	HAM-D$_{17}$/IDS-C$_{30}$	20-25 min	Telephone	ROA	Home/Work
	QIDS$_{16}$**	6 min	Telephone	IVR	Home/Work
Function	SF-12	4 min	Telephone	IVR	Home/Work
	WSAS**	2 min	Telephone	IVR	Home/Work
	WPAI**	2.5 min	Telephone	IVR	Home/Work
Quality of Life	Q-LES-Q	6 min	Telephone	IVR	Home/Work
Side Effects	FISER/GRSEB**	1.5 min	Telephone	IVR	Home/Work
Participant Satisfaction	PSI	1 min	Telephone	IVR	Home/Work
Utilization & Cost	UAC PQ-15	5 min	Telephone	IVR	Home/Work
Income	IPAQ	3 min	Telephone	ROA	Home/Work

Outcome Assessments are obtained at the exit from each treatment level and in follow-up at months 3, 6, 9, and 12.

**Also obtained at week 6 of each level.

ROA = Research Outcomes Assessor
IVR = Interactive Voice Response

FISER/GRSEB = Frequency and Intensity of Side Effects Rating/Global Rating of Side Effects Burden
HAM-D$_{17}$ = Hamilton Rating Scale for Depression (17-item)
IDS-C$_{30}$ = Inventory of Depressive Symptomatology – Clinician-Rated (30-item)
IPAQ = Income and Public Assistance Questionnaire (5-item)
PSI = Patient Satisfaction Inventory (2-item)
Q-LES-Q = Quality of Life Enjoyment and Satisfaction Questionnaire
QIDS$_{16}$ = Quick Inventory of Depressive Symptomatology (16-item)
SF-12 = Short-Form Health Survey (12-item)
UAC PQ-15 = Utilization and Cost Patient Questionnaire (15-item)
WPAI = Work Productivity & Activity Impairment Questionnaire (6-item)
WSAS = Work & Social Adjustment Scale (5-item)

After written informed consent at pretreatment (baseline), participants are evaluated by the CRC. Clinical and demographic information are collected, as well as prior course of illness, current and past substance abuse, prior suicide attempts, family history of MDD or BPD, current general medical illnesses, and prior history of treatment in the current major depressive episode (both medication and psychotherapy) (see below).

Participants complete modified paper and pencil version of the Psychiatric Diagnostic Screening Questionnaire (PDSQ)[245] in the clinical site office. The CRC completes the HAM-D$_{17}$ (at baseline to establish inclusion/exclusion criteria) and the QIDS-C$_{16}$ and reviews inclusion/exclusion criteria. Current general medical conditions (GMCs) will be assessed by the Cumulative Illness Rating Scale (CIRS).[135,157] The 14-item CIRS is completed by the CRC or clinician using a manual[158] to guide scoring, and to gauge the severity/morbidity of GMCs relevant to different organ systems.

The CRC registers the participant using the Interactive Voice Response (IVR) system. After a participant number is assigned, the participant is provided with a medication card and a prescription for a 1-week supply of CIT (20 mg/d qam), which may be started at 10 mg/d x 3 days depending on clinician choice (e.g., in the elderly). The medication card will allow the participant to obtain all study medications at no cost from any pharmacy, throughout the study. The details of dosing and how to switch from one protocol treatment to another is defined in the Clinical Procedures Manual. The participant is informed by the CRC that the ROA will call the participant for a telephone interview within 72 hours of the baseline visit to complete the HAM-D$_{17}$, the IDS-C$_{30}$, and the IPAQ. The participant also completes baseline IVR measures within 72 hours (see Table 1).

The IVR method to acquire outcomes data from clinical trials in outpatients with nonpsychotic mood or anxiety disorders has been the subject of much research over the last decade.[166] Symptomatic outcomes compare favorably with standard clinical interviews in the assessment of anxiety, depressive, and obsessive-compulsive symptoms.[125] Recent reviews (e.g., Kobak et al.[126]) reveal: (1) high subject acceptance, (2) high "interrater" reliability, (3) high correspondence between clinical and IVR interviews, (4) sensitivity to change over time, and (5) equivalent capacity to distinguish symptomatic outcomes between treatment groups when compared to standard clinical ratings. Specifically, four depression studies have compared IVR HAM-D ratings with clinician-completed HAM-D interviews.[4,22,127,128] Furthermore, function, satisfaction, work activity, and side effects can also be gauged using IVR. The STAR*D protocol continues to rely upon interview-based outcomes assessment for symptom severity ratings. However, it will rely upon IVR for all other major outcomes (function, side effect burden, and satisfaction).

I. Research Outcomes Assessments

Recall that research outcomes assessments are distinguished from assessments conducted at clinic visits. The latter are designed to collect information that guides clinicians in the

implementation of the treatment protocol. Research outcomes assessments are not collected at the clinic visits. They are not collected by either clinicians or CRCs.

Research outcomes are collected by the ROAs who complete the HAM-D_{17}, the IDS-C_{30}, and the IPAQ (income questions) using a telephone interview. Other research outcomes are collected by IVR (function, quality of life, side effect burden, and participant satisfaction). IVR is also used to obtain additional participant-reported information to supplement the administrative database information for the utilization and cost analyses (the UAC).

Overview: Within 72 hours of the baseline visit, and within 72 hours of exit from each treatment level, an outcomes assessment package is completed. These research outcomes assessments will, therefore, be completed on the 4,000 participants with MDD at entry into STAR*D and at exit from Level 1; on $\geq$2,000 participants at entry into and at exit from Level 2; and so forth for each subsequent treatment level. (Note: The research outcomes assessment upon exit from one level constitutes the research outcomes .assessment for entry into the next level or for entry into follow-up.)

This same full assessment package is also completed every 3 months for all participants who enter the 1-year naturalistic follow-up from all levels of treatment. A monthly IVR will be used to collect interim follow-up assessments monthly (see below).

The following research outcomes assess several specific key outcome domains. The total package will not exceed 60 minutes: 30 minutes for IVR, and 30 minutes for a telephone interview with the ROA to collect the HAM-D_{17}, the IDS-C_{30}, and the IPAQ.

Symptoms —	The 17-item Hamilton Rating Scale for Depression (HAM-D_{17})[80,81,179,241] (primary) and the 30-item Inventory of Depressive Symptomatology – Clinician-Rated (IDS-C_{30})[198,199] (secondary) are the research outcomes symptom measures. We will collect the IDS-C_{30} and the HAM-D_{17} at entry into and exit from each level, and every 3 months in follow-up. These measures are collected by the ROAs. We will also collect the QID-S_{16} by IVR to determine how this method performs compared to the above two "gold" standards.
Function —	The Short-Form Health Survey (12-item) (SF-12),[232] collected by IVR, is the primary functional outcome. Secondary functional outcome measures are total scores on the (a) 6-item Work Productivity and Activity Impairment Questionnaire (WPAI),[186] and (b) the 5-item Work and Social Adjustment Scale (WSAS)[165] – each collected by IVR.
Quality of Life —	The 16-item Quality of Life Enjoyment and Satisfaction Questionnaire (Q-LES-Q),[45] collected by IVR, will assess quality of life (6 minutes).
Side Effects —	The primary side effect outcome measures are three global ratings: one for side effect frequency; one for side effect intensity; and one

for side effect burden. The FISER (Frequency and Intensity of Side Effect Rating) and GRSEB (Global Rating of Side Effect Burden) include global measures, each using a 7-point Likert-type scale rated 1-6, one rating anchored for frequency, another rating the intensity of side effects encountered in the prior week that the participant believes were due to the antidepressant treatment, and the third asking participants to estimate the overall burden or degree of interference in day-to-day activities and function due to the side effects attributable specifically to the antidepressant treatment. All are collected by IVR.

Participant Satisfaction — A short (2–item) Patient Satisfaction Inventory (PSI) measures satisfaction with the treatment and the treating personnel.

Utilization and Cost — The 15-item Modified Utilization and Cost Patient Questionnaire (UAC-PQ$_{15}$) will gather treatment information from the participant for both mental and general medical conditions in the prior three months.

Income The 5-item Income and Public Assistance Questionnaire (IPAQ) measures the participant's monthly income and the source of the monthly income (e.g., employment wages, public assistance).

I-1.Depressive Symptoms: ROA-rated symptom severity measures will determine symptomatic outcomes. Core depressive symptoms are collected by both measures (the HAM-D$_{17}$ and the IDS-C$_{30}$). Associated symptoms (e.g., atypical symptom features) are collected by the IDS-C$_{30}$.

I-1a.   Clinician Ratings: We have selected the HAM-D$_{17}$ (primary) and the IDS-C$_{30}$ (secondary) as the two major clinical interview ratings of symptom severity. The IDS-C$_{30}$ measures all core symptoms, as well as reverse neurovegetative and other atypical symptom features (e.g., hypersomnia, hyperphagia, interpersonal rejection sensitivity). The HAM-D$_{17}$, the primary outcome, allows comparison to the vast RCT literature. The IDS-C$_{30}$ appears to be more sensitive to change and to residual symptoms than the HAM-D$_{17}$.[12] The ROA uses a semi-structured interview to acquire both HAM-D$_{17}$ and IDS-C$_{30}$ ratings in 20-25 minutes without being repetitious.

Comparing the IDS-C$_{30}$, collected by the ROA, and the QIDS$_{16}$, collected by IVR, allows us to determine the degree to which a briefer itemized symptom rating obtained by IVR can be substituted for a clinician rating. If this briefer rating can substitute for a clinician rating, the dissemination and implementation of STAR*D findings is made easier. Thus, the inclusion of the QIDS$_{16}$ by IVR is aimed at methodological improvements.

I-2.Function: There is no commonly agreed upon gold standard for evaluating function in persons with mental illnesses. We have selected the SF-12 as a primary function outcome measure because it is commonly used in mental health services research and in studies of general medical disorders. The SF-12 has 2 subscales: mental function and

physical function, which closely approximate those obtained with the longer SF-36.[232] The SF-12 is collected by IVR (4 minutes).

As a primary outcome measure for work function, we will use the 6-item WPAI acquired by IVR (2.5 minutes). This measure focuses on work and related function. As a secondary measure of overall work and social function we will use the WSAS (5 items using a Likert-type scale) acquired by IVR (2 minutes).

I-3. Quality of Life: The Q-LES-Q is designed to measure satisfaction and enjoyment, as opposed to function per se, in various domains: physical health, mood, work, household duties, school/course work, leisure time activities, social relations, and general activities. The short version of the Q-LES-Q has 16 items, and is obtained in 6 minutes by IVR.

I-4. Side Effect Burden: Medication side effects are major deterrents to adequate pharmacotherapy, and are often the basis for changing doses or even classes of antidepressants. They often limit doses possible for a particular participant. It is, however, not simply the number, frequency, intensity, nor particular types of side effects, but especially it is the burden of these side effects on participant daily function that affects continuation or discontinuation of the treatment. To compare one treatment with another, we assess the frequency and intensity (2-items), as well as the day-to-day burden (1-item) of the side effects of the antidepressant treatment(s) as defined by the participant (FISER/GRSEB), as part of the research outcomes assessments. Side effect questions are collected by IVR (1.5 minutes).

I-5. Participant Satisfaction: At each research outcomes assessment, participant satisfaction will be measured by 2 Likert-type questions using IVR (the PSI). These questions evaluate satisfaction with the treatment and the clinical staff providing the treatment.

I-6. Acquisition of Interim Research Outcome Data: At the midpoint (6 weeks from entry into each treatment level), we will use the IVR to acquire the $QIDS_{16}$, the WSAS, the WPAI, and the FISER/GRSEB. If the participant has already exited the level, these data are not collected.

I-7. Follow-up Assessments: Once participants remit or respond but decline further randomization, they will be reconsented and may enter follow-up. Follow-up research outcomes assessments are of two types: (1) full assessments conducted every 3 months for a period of 12 months, and (2) interim follow-up assessments conducted monthly on those months when the full research outcomes assessment battery is not obtained.

The full research outcomes assessment battery is identical to the research outcomes assessments completed at entry into and exit from each level of treatment with the exception of one additional measure that assesses whether the participant has sought any medical treatment and, if so, for what and by whom (i.e., ROA telephone interviews to complete the $HAM-D_{17}$, $IDS-C_{30}$, and IPAQ; IVR to obtain function, side effect, and satisfaction ratings; and the $QIDS_{16}$). Interim follow-up assessments are identical to the

interim assessment conducted at week 6 in each level of treatment (i.e., QIDS$_{16}$, WPAI, WSAS, FISER/GRSEB, all by IVR). (See Tables 2 and 3.)

Note that baseline research outcomes assessments match the research outcomes planned thereafter, except that the FISER/GRSEB and the PSI are not obtained at baseline. Furthermore, diagnostic intake information is obtained only at study intake. The clinical visit data pertains to psychotherapy and to medication clinical visits. For participants who receive both medication and psychotherapy, clinical visit measures are collected by both the psychotherapist and clinician prescribing the medications at weeks 0, 2, 4, 6, 9, and 12.

Table 3. Schedule of Evaluations For Follow-Up

Domain	Measure	Time	How	When	Who	Where
Post-Level Follow-up						
Symptoms	HAM-D$_{17}$/IDS-C$_{30}$	20-25 min	Telephone Call	Months 3, 6, 9, 12	ROA	Home/Work
	QIDS$_{16}$	6 min	Telephone	Months 3, 6, 9, 12	IVR	Home/Work
Function	SF-12	4 min	Telephone	Months 3, 6, 9, 12	IVR	Home/Work
	WSAS	2 min	Telephone	Months 3, 6, 9, 12	IVR	Home/Work
	WPAI	2.5 min	Telephone	Months 3, 6, 9, 12	IVR	Home/Work
Quality of Life	Q-LES-Q	6 min	Telephone	Months 3, 6, 9, 12	IVR	Home/Work
Utilization & Cost	UAC PQ-15	5 min	Telephone	Months 3, 6, 9, 12	IVR	Home/Work
Side Effects	FISER/GRSEB	1.5 min	Telephone	Months 3, 6, 9, 12	IVR	Home/Work
Participant Satisfaction	PSI	1 min	Telephone	Months 3, 6, 9, 12	IVR	Home/Work
Income	IPAQ	3 min	Telephone Call	Months 3, 6, 9, 12	ROA	Home/Work
Treatment	TX	2 min	Telephone Call	Months 3, 6, 9, 12	ROA	Home/Work
Post-Level Follow-up Interim Research Evaluations						
Symptoms	QIDS$_{16}$	6 min	Telephone	Months 1,2,4,5, 7,8,10,11	IVR	Home/Work
Function	WSAS	2 min	Telephone	Months 1,2,4,5, 7,8,10,11	IVR	Home/Work
	WPAI	2.5 min	Telephone	Months 1,2,4,5, 7,8,10,11	IVR	Home/Work
Side Effects	FISER/GRSEB	1.5 min	Telephone	Months 1,2,4,5, 7,8,10,11	IVR	Home/Work

IVR = Interactive Voice Response

FISER/GRSEB = Frequency & Intensity of Side Effects Rating/ Global Report of Side Effect Burden
HAM-D$_{17}$ = Hamilton Rating Scale for Depression (17-item)
IDS-C$_{30}$ = Inventory of Depressive Symptomatology – Clinician-Rated) (30-item)
IPAQ = Income and Public Assistance Questionnaire (5-item)
PSI = Patient Satisfaction Inventory
QIDS$_{16}$ = Quick Inventory of Depressive Symptomatology (16-item)
Q-LES-Q = Quality of Life Enjoyment and Satisfaction Questionnaire
SF-12 = Short-Form Health Survey (12-item)
TX = Treatment Questions
UAC PQ-15 = Utilization and Cost Patient Questionnaire (15-item)
WPAI = Work Productivity & Activity Impairment Questionnaire (6-item)
WSAS = Work & Social Adjustment Scale (5-item)

Table 4. Schedule of Measures Acquired at Clinic Visits*

Domain	Measure	Time	How	Who	Where	When
Symptom	CGI-I	1/2 min	Interview (1-item)	CLIN	Clinic	All Visits
	QIDS-C$_{16}$	3 min	Interview (16-item)	CRC/CLIN	Clinic	All Visits
	QIDS-SR$_{16}$	4 min	Self Report (16-item)	Participant	Clinic	All Visits
Side Effects	FISER/GRSEB	1.5 min	Self Report (3-item)	Participant	Clinic	All Visits
	PRISE	3 min	Self Report (8-item)	Participant	Clinic	All Visits
Medication Compliance	MAQ	3 min	Self Report	Participant	Clinic	All Visits

*Note: These measures are used to provide consistent information to the clinicians who use this information in the protocol and are recorded on the CRF.
These measures are collected at clinic visits for participants in protocol treatment. They are not collected at clinic visits for participants seen in follow-up.

CGI-I = Clinical Global Impression — Improvement (based on each level of treatment)

FISER = Frequency & Intensity of Side Effect Rating
GRSEB = Global Rating for Side Effect Burden
MAQ = Medication Adherence Questionnaire
PRISE = Patient Rated Inventory of Side Effects
QIDS-C$_{16}$ = Quick Inventory of Depressive Symptomatology - Clinician-Rated (16-item)
QIDS-SR$_{16}$ = Quick Inventory of Depressive Symptomatology - Self-Report (16-item)

J. Clinic Visits

J-1. Clinic Visit Assessments: Measurements are collected at clinic visits (i.e., when participants are seen for treatment purposes) at all protocol levels. Clinic visit data are not collected for participants in follow-up. Measurements acquired at clinic visits provide a consistent information base for all clinicians so that they can implement the protocol consistently. These measurements include the QIDS-C$_{16}$ (CRC), the QIDS-SR$_{16}$ (participant self-report), the FISER/GRSEB, and the PRISE (all participant self-reports). At each clinical visit, clinicians also complete the Clinician Global Impression – Improvement (CGI-I) scale.

Table 5 shows the synopsis of data collection at each treatment level, as well as the measures acquired at each clinical visit. The research outcomes are collected at exit from each treatment level (e.g., Levels 1, 2, 2A, 3, 4, etc.). Research outcomes are also collected at months 3, 6, 9, and 12 in the one-year, naturalistic follow-up for participants with a satisfactory therapeutic response to any level of treatment.

J-2. Adverse Events: Adverse events are recorded by the study clinicians or CRCs at each clinic visit, using both spontaneous reports and participant responses to the PRISE, an 8-item modified version of the Somatic Symptoms Scale (SSS). The SSS differentiates the side effects of different classes of antidepressants.[224] We have modified the SSS to include common side effects of the SSRIs, TCAs, MAOIs, MIRT, BUP, Li, and other agents in the STAR*D algorithm. Individual symptom clusters are rated on a 3-point scale (0: absent, 1: mild or intermittent, 2: severe or persistent). The PRISE should take 3 minutes for participants to complete. At these clinic visits, we also obtain the FISER/GRSEB to inform the providers who must make dose adjustment and other treatment decisions.

A Serious Adverse Event (SAE) is any adverse drug experience occurring during the study that: (1) results in death, (2) is life-threatening, (3) results in hospitalization or prolongation of hospitalization, (4) results in persistent disability, or (5) results in congenital anomaly. Important medical events that may not result in death, be life-threatening, or require hospitalization may also be considered SAEs if medical judgement suggests that, in the absence of medical intervention, one of the above conditions would ensue or otherwise jeopardize the participant.

SAEs are to be reported both verbally and in writing (SAE Reporting Form) to the RC Director (or his/her physician designee in case of absence). The RC Director must consult with the study clinician to determine whether to taper or discontinue the participant's study medication(s).

The RC Director or his/her physician designee must be available 24 hours a day/7 days a week to respond to the reporting of serious adverse events. They, in turn, will make the determination whether the event is related to the study medication(s) and then report it to the Co-Director, NCC, and STAR*D Safety Officer. These reports must be completed within 24 hours of the occurrence of the serious adverse event. The STAR*D Safety Officer is available 24 hours a day/7 days a week to respond to reports of serious adverse events or any other safety issues. The Safety Officer, with the assistance of the Clinical Manager, reports these events to the STAR*D Principal Investigator (PI) and CoPI and to the Data and Safety Monitoring Board (DSMB), the RC Institutional Review Board (IRB), the Food and Drug Administration, and to the manufacturers of the medication(s) involved in the reports. The same reporting requirements apply to all other events the study clinician deems serious, regardless of whether or not they meet the above criteria.

Table 5. Overview of Clinical and Research Data Collection at Each Treatment Level

Data	How	Intake	Weeks: 2	4	6	9	12 (or exit)
Intake							
Consent	(CRC)	√					
Clinical/Demo	(CRC)	√					
Incl./Excl.	(CRC)	√					
PDSQ	(Participant)	√					
HAM-D$_{17}$/QIDS-C$_{16}$	(CRC)	√					
CIRS	(CRC)	√					
Measures at Clinic Visits							
QIDS-C$_{16}$	(CRC)	√	√	√	√	√	√
QIDS-SR$_{16}$	(Participant)	√	√	√	√	√	√
CGI-I	CLIN	NA	√	√	√	√	√
PRISE	(Participant)	NA	√	√	√	√	√
FISER/GRSEB	(Participant)	NA	√	√	√	√	√
Research Outcomes Assessments							
HAM-D$_{17}$	(ROA)	√					√
IDS-C$_{30}$	(ROA)	√					√
IPAQ	(ROA)	√					√
QIDS$_{16}$	(IVR)	√			√		√
SF-12	(IVR)	√					√
WSAS	(IVR)	√			√		√
WPAI	(IVR)	√			√		√
Q-LES-Q	(IVR)	√					√
FISER/GRSEB	(IVR)				√		√
PSI	(IVR)						√
UAC-PQ$_{15}$	(IVR)	√					√

K. Cost/Utilization

K-1. Direct Costs and Indirect Costs: The main focus of STAR*D is to define empirically the outcomes of various next treatments for participants with less than acceptable clinical responses to prior treatment(s). STAR*D, however, provides an opportunity to gather prospective information on participant use of both medical and mental health services, and the cost of those services, or the "direct" cost of health care.

Components of direct care costs include (for STAR*D): (1) the planned program intervention including staff time; (2) other psychiatric care for Major Depressive Disorder; (3) other mental health and addiction care services participants seek for other mental health problems; and (4) general medical care.

The protocol also affords investigators the opportunity to collect information about the "indirect" cost of major depressive disorder. Indirect costs are described in the "cost of illness" literature[188,189,190,219] that detail the economic "burden" of a disorder to society.

These economic components include: **(A)** *lost productivity* (including market-based employment and household production) due to days lost due to morbidity and premature mortality and reduced productivity, **(B)** *administrative costs* (including administration of welfare transfer payments, caregiving time, and criminal justice such as police, courts, jails, and prisons), and **(C)** *consequential damages* (including lost time due to incarceration, property loss, and suffering of third party victims such as family member experiences and death of family member due to suicide[112]). Thus, changes in health outcomes can be translated into these economic components.

K-2. Focus on Direct Costs: Our emphasis on economic costs will be on measuring the direct cost of care as part of an overall strategy to compare "outcomes" with costs.

Costs are important factors in evaluating clinical practice guidelines, [131,137,149] where the goal is to apply the latest medical knowledge to optimize participant health outcomes (outcome driven) or to achieve the biggest "bang for the buck" (value driven). The latter requires balancing health outcomes against costs. Thus, it is appropriate to use a cost-outcome framework to evaluate guideline-driven clinical programs.[110]

Frameworks for cost-outcome studies vary in the literature. Between 1991 and 1996, the medical literature contained some 3,500 cost-outcome publications,[43] with methods cited in economics[160] and human resources;[86,142] environmental regulation;[9] pharmacy;[161] clinical decision making[40] and management;[243] public health[26] including health services,[234] medical care quality,[34] and program evaluation;[37] and clinical trials.[212] Cost-outcome methods have also been summarized by medical specialty, including medicine,[94,237] oncology,[209] cardiology,[130] and mental health.[83,159,240] These studies vary with respect to (*i*) how outcomes are defined, (*ii*) what costs are included, and (*iii*) how costs and outcomes are compared.

Our study focuses on how the direct costs of the treatment intervention act as an input into the production of a health care outcome. Thus, we focus on estimating direct costs of the participant's care. This is consistent with the U.S. Public Health Service (PHS) sponsored Panel on Cost-Effectiveness in Health and Medicine.[73,205,213,236] Considered as program outcomes, changes in indirect costs are treated on the "benefits" side of the cost-outcome scale, and thus are excluded from the cost calculus to avoid double counting.[115]

Our focus on the cost of both mental and general medical services, as recommended by the PHS Panel, is particularly compelling in light of empirical studies evidencing an association between use of general medical and mental health care, and underscores Simon and Katzelnick's comment that cost offsets of depression will need to be visited by "…a new generation of experimental studies adapted to assess economic outcomes."[214] In fact, the issue of cost offsets in the treatment of depression has not been resolved. Lave et al.'s randomized controlled trial of "ambulatory medical populations treated in routine practice…"[133] who meet criteria for major depression, found no offsets when comparing interpersonal psychotherapy and nortriptyline pharmacotherapy with "usual care." Such findings are echoed in Zhang et al.'s analysis of community residents in a descriptive analysis that found no difference in physical care costs for participants receiving

treatment for depression by a mental health specialist rather than a primary care provider.[244] On the other hand, Guthrie et al. found cost offsets compared to controls for treatment nonresponders randomized to psychodynamic-interpersonal therapy.[77] There are several bases for finding a cost offset.

As reviewed by the investigators,[109,111] patients frequently seek general medical providers to care for psychiatric problems (substitution effect).[30,41,168] Second, the presence of psychiatric symptoms tend to be associated with more medical care costs (health effect).[14,103,121,134,141,153,215] Third, patients in clinical environments who use specialty mental health care tend to use fewer general medical care services, including outpatient visits,[177] inpatient days,[39] and peak use of primary care services,[82] and at lower general medical costs (cost-offsets).[95,104] Fourth, clinical observers report reductions in use of general medical care following use of mental health care,[90,164,195] treatment for undiagnosed panic disorder,[206] and psychotherapy,[29] while others have reported no change,[215] or an increase[152] in general medical care following use of mental health services (discovery effect). Finally, the effectiveness of general medical care may be influenced by psychiatric status (productivity effect).[68] For example, psychiatric status has been observed to be related to mortality rates among the elderly with hip fractures,[140,162] and myocardial infarction.[66] Finally, economists generally prefer to report costs in terms of a total burden on society. While important for social policy decisions, such information would do little to motivate individual providers to adhere to, participants to comply with, or third parties to finance care based on, clinical guidelines. Thus, we calculate payer-specific costs to describe how the cost burden is shared among participants, third party payers, and the health care providers.

K-2.a   Computing Direct Costs of Care: Costs of care are computed by multiplying units of use as defined by a common service classification system, by a unit cost for each health service type, and summing over all services.[73,83,111,154]

K-2.b   Services Classified into Common Service Categories: For our purposes, services are classified into common categories that include (1) outpatient procedure by CPT code, (2) inpatient day by DRG, (3) acute psychiatry inpatient days, (4) long-term inpatient psychiatry, (5) nursing home or extended care facilities; and (6) Prescription Drugs by National Drug Code.

K-2.c   Unit Costs: Unit costs are computed for each service category based on an existing schedule of rates. Sources for these schedules include actual charges from participating clinical site providers and third party payers, as well as secondary sources including Department of Veterans Affairs cost accounting system (Cost Distribution Report, Decision Support System) and community charge based schedule (VA's Medical Care Cost Fund Community Provider Charge Schedule derived from Milliman and Robertson's National Physician Charge Databases), Health Care Financing Administration fee schedule (Physician Fee Schedule and DRG Reimbursement under Medicare), and state-level Medicaid UCR fee schedules. Differences in cost estimates by unit cost schedules are compared using sensitivity analyses. In the absence of cost-based prices in health care markets, this strategy enables investigators to determine how much

participant care would cost if treatment were set in different cost environments with different payers (Medicaid bills, VA medical centers, Medicare reimbursement rates, provider charges and collections, and third party payer rates).

K-3. Use of care: Use of care will be constructed from clinical site-specific administrative databases and from participant self-reports that, when taken together, will be used to construct a "hybrid"[25,108] or a "best estimate"[220] of participant use of care.

K-3.a Provider-based Records: Provider records have always been the gold standard to measure service use.[114] We measure use of care by accessing records from participating providers at each clinical site. Part of the criteria to select clinical sites is the availability and quality (accuracy, completeness) of such records.

File sources include, among others, billing and claims files from health care providers, payment records from third party payers, accounting databases from government facilities, and utilization review from managed care services. Since data sources vary in terms of variable labels and field definitions, service codes from selected clinical sites will be "reconfigured" to bridge utilization information into common service categories necessary to compute "costs." See section on "unit costs" above. These constructions will be done by consensus conference on a site-by-site basis. The conference will include selected members of the clinical site's medical staff, data manager, STAR*D study staff and investigators.

K-3.b Participant Self-Reports (UAC-IVR): When provider-based records are inaccessible, information collected by interviewing participants will serve as an adequate substitute. Estimates of participant use of both mental and general health care for the entire sample will be computed from participant responses to the Utilization and Cost Methodology Interactive Voice Response system (UAC-IVR). This system was derived from the Utilization and Cost Patient Questionnaire[113] (UAC-PQ) as applied in the Texas Medication Algorithm Project.[111] Participants will be asked to dial a phone number and record their use of care information by means of a touch tone or voice activated telephone. Structured questions will ask participants if they have visited a medical doctor, psychologist, or social worker in an outpatient care clinic during the past three months for their depressive disorder, for other mental health conditions, for addiction disorders, and for general medical care. If so, participants are asked to describe how many visits they spent in each category. Participants are also asked about visits to emergency rooms and days spent in the hospital. Of course, whenever possible, participant utilization information will be supplemented by provider-based files.

It is impractical to expect participants to recall service use beyond simple "visits" and "days." This is a problem because unit costs, required to calculate direct costs of care, are measured in terms of common service categories (per diem by DRG, procedure by CPT code, etc.). Thus, participant responses must be translated directly into a dollar value.

Such costs will be computed from participant responses with the aide of a mathematical model. The model will be constructed, estimated, and tested on study participants who

completed both the UAC-IVR and for which provider-based information is also available to compute direct health care costs. The dependent variable is direct health care costs computed for each unit cost schedule. Independent variables include the UAC-IVR participant reported units of use (visits, days) and care type (general medical or psychiatric). To adjust for participant-specific confounds with use of care, covariates include participant characteristics reflecting enabling (monthly income, own home, family and household size), predisposing (education), and need (baseline outcome assessments, other comorbid conditions) factors. The form of the model will be derived as a "best fit" Taylor series expansion of first order, interaction terms, and higher ordered terms derived using backward stepwise regression. This method is being applied in the Texas Medication Algorithm Program.[111]

K-4. Analysis of Cost Data: To compare two treatment options with respect to direct health care costs in a repeated measures, multi-site, participant randomized, and panel-design longitudinal study, we employ a growth curve analysis.[217,242] Also known as random-effects,[132,233] random regression,[71,97] empirical Bayes,[144] general mixed linear,[74] and hierarchical linear[20] models, the procedure allows investigators to estimate effect size and test for significance using t-tests. Computations will be based on HLM/3L software.[19]

Our analytic approach takes into account repeated measures and randomization by participant; it does not require fixed intervals between actual follow-up observations; it can measure how effects vary with time; and it allows for more flexible covariance structures for a better model fit.

Specifically, to compare, say, treatment option "1" with an alternative treatment option, consider:

y_{it}	=	costs for participant "i" at time "t."
β_0	=	costs at baseline for participants with alternative treatment.
$\beta_0 + \beta_1$	=	costs at baseline for treatment option 1 participants.
β_1	=	difference in costs at baseline between treatment option 1 and alternative treatment participants, if any.
β_2	=	rate of change in costs over time for alternative treatment participants, or a time trend.
$\beta_2 + \beta_4$	=	rate of change in costs over time for treatment option 1 participants, or a time trend.
β_4	=	difference in rate of change in costs over time between treatment option 1 and alternative treatment participants.
β_3	=	rate of change in the rate of change in costs over time for alternative treatment participants.
$\beta_3 + \beta_5$	=	rate of change in the rate of change in costs over time for treatment option 1 participants.
β_5	=	difference in rate of changes in the rate of change in costs between treatment option 1 and alternative treatment participants.
t	=	time measured since baseline ($t=0$ at baseline).

I_i =   dummy variable assigned "1" for treatment option 1, and "0" for alternative treatment participants.

v_i =   random effect for participant i, assumed to be independent, normal random variates.

u_{it} =   random effect for participant i at time t, assumed to be independent, normal random variates and independent of v_i.

Thus, in a randomized participant design, we estimate:

$$y_{it} = \beta_0 + \beta_1[I_i] + \beta_2 t + \beta_3 t^2 + \beta_4[tI_i] + \beta_5 t^2 I_i + v_i + u_{it} \qquad \text{Eq. 1}$$

where treatment effects are represented as a growth rate (β_4) and a change in that growth rate (β_5) over time. Actual differences in costs between treatment options can be computed by calculating the difference in the estimated equations between participants over the study period.

These analyses may be expanded in several ways, with the appropriate model selected to have the best overall fit of the data. These include expanding level-1 to include higher orders of time (t^3, t^4, etc.) to account for non-linear changes in the growth rate with time. Other expansions, including inverse (t^{-1}) and exponential (e^{-at}) will also be explored, and the model with the best fit of the data used to report effect sizes. The random effects may be expanded to include regression to the mean,[10] heteroscedastic and autocorrelated level-1 co-variance structures,[24,88,136] and censored samples that include non-users of services.

The impact of non-users of services on costs is worthy of special note. The issue becomes important when a few participants use expensive services (e.g., inpatient care), causing costs to be bimodally distributed with a long right tail. The issue has been explored by Duan,[38] extended in Pohlmeier and Ulrich,[181] and described econometrically in Maddala.[138] For this study, use of expensive, though rare items will be explored separately in which use versus non-use (logistic regression) is modeled separately from the volume of use among care users (linear regression). For the logistic regression, equation 1 would be re-configured so that the dependent variable, or "y_{it}," would assume the value of one if participant i consumed the rarely used service at time t, and zero otherwise. For the volume variable, the long right tail in the cost variable can be managed by transforming costs using a log, or other appropriate, function.[2,124,227]

The investigators are also prepared to compute the impact on health care costs in cases where participants were not randomized into treatment. In such cases, the treatment group indicator variable (I_i) would be measured using an instrumental variable constructed from predicted values, a method well-known in econometrics[85,86,87] and suggested for mental health cost-outcomes research.[83]

Finally, equation 1 may also be modified to include measures of participant health care needs (psychiatric diagnosis, other medical conditions), predisposition to care (education, age, ethnicity, participant perception of benefits), and access to care (income, disposable income, existence of primary care provider, time (minutes) to obtain care, availability of

health insurance coverage). Exclusion of these relevant factors to predict costs in randomized designs should not bias the estimated impact of treatment on costs. However, model mis-specification may impact significance tests[24,88] and so these other factors will be explored by expanding equation 1.

Finally, comparing costs with changes in outcomes can be done with the aide of Table 6. Here, option 1 compared to alternative treatment is clearly preferred whenever outcomes are favorable, and costs are equivalent, or less than, alternative treatment. Option 1 is also preferred whenever outcomes are equivalent, but costs are less.

Table 6. Policy Choices Between Option 1 and Alternative Treatment Strategies

		COSTS		
		higher	no difference	lower
	Better	*unclear*	Option 1	Option 1
OUTCOME	no change	Alternative treatment	*equivalent*	Option 1
	Worse	Alternative treatment	Alternative treatment	*unclear*

A more common results is that both outcomes improve and costs increase. In this case, it is unclear if treatment option 1 represents a better social choice. To aide in this decision, we intend to compute cost-outcome ratios based on two formats.[110] The cost-outcome "ratio of differences" statistic is based on Jerrell and Hu[98] formulation and will measure how much additional outcome is produced for each additional health care dollar that participants in treatment option 1 incur over alternative treatment participants. For randomly assigned participants, the statistic is computed by:

$$Cost\text{-}outcome\ Ratio\ of\ Differences = \frac{\bar{o}_T - \bar{o}_C}{\bar{c}_T - \bar{c}_C}, \quad Eq.2A$$

where $\bar{o}_T$ and $\bar{o}_C$ are average outcomes and $\bar{c}_T$ and $\bar{c}_C$ are average costs for treatment option 1 and alternative treatments, respectively. This statistic is intended to help policy makers balance option 1's better health outcomes with anticipated increases in treatment costs.

Better outcomes from treatment option 1 may be the result of more care, rather than from more efficacious treatment. The cost-outcome "difference in ratios" statistic measures the extent, if any, to which an option will yield a greater outcome, dollar for dollar, than an alternative. Calculations are based on Siegel[212] revitalization of Grossman's health production model[76] in which health care costs are treated as inputs (independent variable) to produce health outcomes as outputs (dependent variable). Let $o_T(c)$ and $o_C(c)$ be mathematical functions describing the relationship between health care costs "c" for participants assigned to option 1 and alternatives, respectively, on health outcomes "o." The term "marginal productivity" refers to the change in outcome associated with an addiction dollar invested in health care under a given option. At a given initial investment in health care of c_o, marginal productivity may be computed as: $[o(c_0 + \$1) - o(c_0)]/\$1$.

The "difference in ratios" cost-outcome statistic equals the difference in marginal productivity between options when both groups have received an equivalent dollar investment in health care, or:

$$Cost\text{-}Outcome\ Difference\ in\ Ratios = \frac{[o_A(c_0 + \$1) - o_A(c_0)]}{\$1} - \frac{[o_U(c_0 + \$1) - o_U(c_0)]}{\$1} \quad \text{Eq. 2B}$$

The statistical significance of these ratios can be computed from an estimated distribution of ratio estimates computed by bootstrapping the sample (sampling from the sample with replacement).

Finally, while the investigators appreciate the importance to measure cost-effectiveness ratios based on *utility*[176] (health states weighted by consumer preferences) or Quality Adjusted Life Years, or *QALYs*[229] (life years adjusted for morbidity due to illness), no plans to compute such estimates are being prepared at this time.
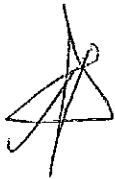
L. Project Overview: Participant Flow

STAR*D is planning to enroll at least 4,000 participants into protocol-defined treatment over a 24-month enrollment period ▮▮▮▮▮▮▮▮▮▮▮▮ Each of 14 RCs will oversee at least 2 but not more than 3 clinical sites. Each RC will activate 2 clinical sites. Depending on participant recruitment, a third clinical site may be activated. Thus, there will be 28-42 clinical sites altogether. If there are 28 clinical sites (2/RC), then each clinical site enters 6.0 participants /month over the 24-month enrollment period. If there are 3 clinical sites/RC (i.e., 42 clinical sites altogether), then each clinical site enters 4.0 participants per month.

Table 7 shows the overall participant flow for the entire STAR*D project (n=4,000 entrants into Level 1 protocol care) as well as the same figures for a single RC (assuming 14 RCs). It also shows the same figures for a clinical site if there are 42 clinical sites, and for a clinical site if there are 28 clinical sites.

These figures are based on the following assumptions:

1) Enrollment Period: 24 months
2) Regional Centers #: 14
3) Treatment Visits at 0, 2, 4, 6, 9, and 12 weeks for medication or psychotherapy
4) Follow-up: 12 months after exit from each treatment level only for participants with a satisfactory therapeutic response
5) Satisfactory Response Rates at each level are as follows:

   Level 1 — 45%
   Level 2 — 40%
   Level 3 — 30%
   Level 4 — 20%

Table 7. Overview of Participant Flow for the STAR*D Project

Entire Project

Level	Entry/ Month	# Active Participants at Peak	# Participants to FU/Month	# Participants to Next Level/Month	Length of Peak Flow (weeks)
1	167	462	75	83	93
2	83	231	33	36	93
In therapy:	18	49	7	7	93
2A	7	20	3	3	93
3	29	80	9	12	81*
4	12	34	3	NA	81*

Per Regional Center (RC)

Level	Entry/ Month	# Active Participants at Peak	# Participants to FU/Month	# Participants to Next Level/Month	Length of Peak Flow (weeks)
1	11.9	33.0	5.4	6.0	93
2	6.0	16.5	2.4	2.6	93
In therapy:	1.3	3.5	0.5	0.5	93
2A	0.5	1.4	0.2	0.2	93
3	2.1	5.7	0.6	0.9	81*
4	0.9	2.4	0.2	NA	81*

Per Clinical Site (if 42 clinical sites, 3 per RC):

Level	Entry/ Month	# Active Participants at Peak	# Participants to FU/Month	# Participants to Next Level/Month	Length of Peak Flow (weeks)
1	4.0	11.0	1.8	2.0	93
2	2.0	5.5	0.8	0.9	93
In therapy:	0.4	1.2	0.2	0.2	93
2A	0.2	0.5	0.1	0.1	93
3	0.7	1.9	0.2	0.3	81*
4	0.3	0.8	0.1	NA	81*

Per Clinical Site (if 28 clinical sites, 2 per RC):

Level	Entry/ Month	# Active Participants at Peak	# Participants to FU/Month	# Participants to Next Level/Month	Length of Peak Flow (weeks)
1	6.0	16.5	2.7	3.0	93
2	3.0	8.2	1.2	1.3	93
In therapy:	0.6	1.8	0.3	0.3	93
2A	0.3	0.7	0.1	0.1	93
3	1.0	2.9	0.3	0.4	81*
4	0.4	1.2	0.1	NA	81*

* Due to Level 2A entering Level 3 12 weeks after those entering from Level 2.

* Numbers shown for "In therapy" represent the subset of participants (within the total as shown for Level 2) that are receiving psychotherapy.

Table 8 estimates the flow of CT participants and associated CT sessions for the entire (n=4,000) STAR*D project, for each RC (assume 14 RCs), for a clinical site (assume 42 clinical sites), and for a clinical site (assume 28 clinical sites).

Table 8. Estimated Number of Participants in Cognitive Therapy

Estimated Number in Cognitive Therapy

	Overall STAR*D	Per RC	Per CS (42 CSs)	Per CS (28 CSs)
Number of participants into CT (switch into) =	204	14.6	4.9	7.3
Number of participants into CIT+CT (augment) =	224	16.0	5.3	8.0
Total =	428	30.6	10.2	15.3
Total number of sessions at 16/participant =	6,847	489.1	163.0	244.5
Sessions/month over 24 months =	285.29	20.4	6.8	10.2

CS = Clinical Site

M. Clinical Laboratory Monitoring

Neither clinicians nor participants are masked to treatments. Therefore, we will use the routine clinical laboratory support available at each clinical site. Laboratory tests to be conducted are selected by the participants' clinicians and individualized to each participant's particular medical situation. Prior to any decision about clinical laboratory monitoring, a medical review of systems and a physical exam will be conducted at the outset on all participants, in accordance with primary and specialty care practice guidelines. As recommended by the Depression Guideline Panel,[32] clinicians will be encouraged to conduct only a limited number of laboratory tests to detect potential general medical causes for the depression unless specific risk factors, specific or unusual symptoms elicited by a medical review of systems, or an atypical course of illness are present. In such cases, the clinician will select additional tests to answer specific diagnostic questions.

The basic laboratory battery will generally include CBC+differential, urinalysis and microscopic, CHEM-16, and a thyroid function test (i.e., TSH, T3, and T4), based on clinician decision in accordance with the Depression Guideline Panel.[32] These tests will be required to have been conducted within 6 months prior to entering into the study. Electrocardiograms before and after treatment with NTP are optional. The results of these laboratory tests will be recorded in the participant's medical record, but not in the CRF, save for their being part of the inclusion/exclusion criteria. If clinically significant laboratory test abnormalities occur, they will be reviewed by the clinician and appropriate further evaluation and treatment will be provided through routine care.

Laboratory tests to monitor blood levels of relevant psychotropic agents (e.g., Li, NTP) will be used as called for by the treating clinicians. Levels are measured within 10–12

hours after the last oral dose at a frequency determined by the clinician until a therapeutic blood level is achieved.

N. Plans to Minimize Dropouts

Depressed participants face numerous obstacles to study participation including feelings of hopelessness, psychosocial stressors, and misgivings concerning the adequacy with which a "protocol" can address their unique clinical circumstances. Among treatment-resistant populations, the additional discouragement, skepticism, frustration, and symptom evolution that may develop across one or more treatment failures present particular challenges to participant retention.

We recognize that a tension exists between (1) the goal of participant retention and (2) the goal of maximizing the generalizability of the proposed study to heterogeneous clinical settings, including those in which the resources for keeping participants in treatment are limited. In considering this tension, this protocol has emphasized ensuring that participants and clinicians make every reasonable effort to implement the treatments properly because the main objective is a comparison of different treatment strategies, substrategies, and options. It is of greater public health significance to compare well-implemented as opposed to poorly-implemented treatments.

On the other hand, in the interests of representativeness and generalization, heroic or extreme efforts at both ensuring fidelity by clinicians and participation by participants cannot be implemented during the study (e.g., large monetary reimbursements to participants for attending each visit). Further, since there will be a diversity of settings (primary and specialty clinicians in public and private care settings), it is very likely that there will be a wide variation in clinical procedures used to attend to participant adherence and attrition (e.g., participant education, following up on missed appointments). This wide variation runs the risk of creating so much variation in adherence and attrition that comparisons of effectiveness may not be internally valid.

Thus, implementing a set of reasonable clinical procedures to minimize participant dropout is critical to the success of this study as it relies on good participant retention across the sequential treatment levels— in some cases extending over many months.

Thus, we will develop and implement a program to train clinical staff to provide to participants (and family members or significant others) specific information which is consistent across clinical sites, providers, and treatment levels. This educational program is consistent with the practice guideline recommendations.[3,32,33] In general, it provides a feasible, inexpensive set of consistent interactions, thereby reducing unneeded (and potentially scientifically disadvantageous) procedural variations across sites and clinicians that could lead to differential attrition.

There are two types of nonadherence that could interfere with the conduct of this trial: (1) treatment nonadherence (e.g., not attending clinic visits or not taking the treatment

medication), and (2) research nonadherence (e.g., not participating in the research outcomes assessments). The following discussion addresses both types of nonadherence.

Ideally, participants who do not remit with the initial treatment will wish to proceed into one or more subsequent treatment levels until satisfactory benefit has been attained. Thereafter, ideally they will then be willing to participate in follow-up assessments aimed at further characterizing the extent and durability of those responses. To encourage sustained study participation, we intend to pursue several strategies for all participants: a) promoting study affiliation; b) educating participants and families about depression and its treatment; c) ensuring timely follow-up for and re-scheduling of missed appointments; (d) compensating participants for the time/effort to participate in the research outcomes assessments (e.g., 30 minutes with a telephone interview with the Research Outcomes Assessor and 30 minutes with the IVR at study entry, at exit from each treatment level, and every 3 months in follow-up); and (e) providing "reminders" (e.g., letters) to alert participants to assessments.

a)   Study Affiliation: Participating participants will be provided printed brochures and a brief, informational videotape ("Introduction to STAR*D") outlining the overarching rationale, aims, and procedures entailed in STAR*D. It will include all relevant aspects of study design and a description of participating centers, and will emphasize the public health significance of STAR*D and the critical role played by the participants. All participant information materials will be approved by the IRB at each RC.

In addition, participants will receive a bimonthly newsletter (the STAR*D Gram) to sustain interest. It will feature new information on depression, and selected updates on STAR*D and on any new ancillary studies such as family-genetic studies. It will feature members at each RC as well. We will limit the scope of these materials so that they provide information relevant to study participation but do not serve a broad didactic role about particular depression treatments that might overlap with particular treatment options. These introductory materials and the newsletter will be produced under the auspices of the Communications Committee.

b)   Education: There are several elements that are known to enhance participant participation in and adherence to treatment. First and foremost is education (of both participants and significant others[7]). In nearly every randomized trial comparing educated efforts versus none (or usual care), the educated group had either greater adherence to the treatment procedures (whether measured by appointment keeping, medication counts, or blood levels) or better outcomes or both. For this reason, participant/family education is recommended as part of the general clinical management of depressed participants in both primary[32,33] and specialty[3] care.

A multistep educational package will be used. This educational package will be provided to all participants.

c)   Follow-up: Timely follow-up and rescheduling of missed clinic visits is crucial in assuring adequate treatment and in minimizing dropouts. A schedule of visits will be

supplied to the CRCs by the EDC as a participant enters a level or enters follow-up. A centralized system for tracking study appointments will be set-up at each RC (or clinical site) as well as a database with contact information for each participant — which will be updated at least every 3 months. The clinic staff will contact participants on the day of the missed appointment, and again within 24 hours if there is no response. A letter signed by the participant's clinician will be sent out within 48 hours of the missed appointment if phone contact has not been established. This letter will encourage prompt rescheduling of the missed appointment, and provide phone numbers for the clinicians. It will also ensure that participants are aware of other local mental health resources in case of emergencies.

This level of follow-up for clinic visits is likely to be more aggressive than usual care offered in some clinical settings. However, we believe that it is justified to ensure participant retention, and to ensure the safety of depressed participants whose failure to keep an appointment may signal clinical deterioration and emergent safety concerns.

d) <u>Compensation</u>: Because many participants may suffer economic hardship by allocating time to the research outcomes assessments — especially hourly wage earners — and because of our and others' experiences with reasonable compensation for each research assessment, we believe $25/assessment is reasonable (1 hour of participant time) ($15 for each ROA and $10 for each IVR assessment).

e) <u>Reminders</u>: Some clinical sites/or RC will already be using a letter or telephone reminder for clinic visits. Others will not. If missed clinic visits become significant at particular clinical sites (e.g., >15%), we will implement a letter reminder system before the next appointment. We will use a letter reminder system for all research outcomes assessments calls.

f) <u>Guidelines for Discontinuing Participants From the Randomized Treatment Study</u>:
   1) Participant request;
   2) Clinicians decide that to discontinue the study is in the best interest of the participant;
   3) Participant becomes unable to attend clinic visits with STAR*D clinicians, thereby precluding further clinical implementation of the protocol; and
   4) Participant becomes unable to participate in research outcomes assessments.

IV. Data Analysis and Power

General Analytic Approach

Confidentiality will be maintained throughout the course of the study. The study database will not include any information that can be linked to the identity of the participant (e.g., name, telephone number, social security number). A unique study identification number will be used as a unique identifier for each participant. This unique identification number will stay with the participant through the course of the study, regardless of the level of treatment.

Describing the characteristics of the study population will begin the data analysis at each level. Descriptive statistics, including measures of central tendency (means, medians, other percentiles) and dispersion (standard deviations, ranges) will be computed for continuous data such as age. Frequency distributions will be estimated for categorical data such as sex. Graphical displays including histograms and box plots will be produced.

To ensure that the participants are representative of the eligible population, the appropriate statistics (e.g., chi-square tests, t-tests, analysis of variance, Wilcoxon) will be used to compare the baseline data between those who were eligible to participate in the study, but refused entry, and those that entered the study. Similar statistical analysis techniques will be used to determine if the randomization scheme provided a balanced sample by comparing the baseline characteristics across treatments.

Every effort will be made to minimize the amount of missing data and participants lost to follow-up. Participants lost to follow-up could possibly bias the data since they may differ from those participants not lost to follow-up. Comparison of baseline factors such as age and education will be made to determine if the participants lost to follow-up differ from those participants not lost to follow-up.

In general there are three types of hypotheses to be investigated: acceptability hypotheses, effectiveness hypotheses, and predictor hypotheses. We will now layout the general analytic approach for each of the types of hypotheses.

The rates of acceptability of a treatment or set of treatment options will be estimated for the acceptability hypotheses. After the proportions for the treatment acceptability have been calculated, confidence intervals for each of the proportions will be calculated. For example, in the analysis of specific Aim D-2 (1), which is designed to estimate the acceptability of the treatment options available in Level 3, we will calculate the proportion (and confidence intervals associated with the proportion) of participants willing to accept each of the treatment strategies.

A number of analytic approaches will be utilized to assess the two remaining types of hypotheses, effectiveness hypotheses and predictor hypotheses. Specifically, chi-square tests for association, logistic regression and survival analysis (or time-to-event analyses), both Kaplan-Meier curves and Cox Proportional Hazards Modeling will be utilized to test these hypotheses. In addition, approaches such as analysis of variance (ANOVA), analysis of

covariance (ANCOVA) and linear regression may also be utilized. The following are descriptions of each of the general analytic approaches along with an example as to how it would be utilized to test a specific hypothesis.

A chi-square test for association will be used to test for bivariate associations between discrete variables of interest (e.g., treatment) and the discrete outcomes of interest (e.g., response). For example, Specific Aim D-2 (4) is designed to identify baseline predictors of intolerance to treatment options at Level 3. For these analyses, a chi-square test will be used to test for association between categorical baseline predictors (e.g., sex) and intolerance.

Logistic regression will be used in a number of instances. One will be to test for bivariate associations between continuous variables (e.g., age) and the dichotomous outcomes of interest. So, in the example provided above for Specific Aim D-2 (4), logistic regression analysis will be used to test for association between the categorical baseline predictors (e.g., age, baseline severity of depression, level of response) and intolerance.

In addition to the bivariate approach, logistic regression will be used to determine if there is an independent effect after controlling for possible confounding variables. For example, Specific Aim D-1 is designed to determine the effectiveness of the four treatments at Level 3. Multiple logistic regression models will be used to determine if there is an independent effect of the treatment after controlling for possible confounding variables. Logistic regression analysis will also be used for exploratory. In these instances a backward stepping or a stepwise approach will be used to identify variables independently associated with the outcome of interest. For example, Specific Aim D-2 (4) is to identify baseline predictors associated with intolerance of the initial treatment. A list of possible variables contributing to the intolerance of the treatment will be identified and entered into the logistic regression model. A selection process, either backward stepping or stepwise will be used to identify factors independently associated with intolerance to the initial treatment.

For all logistic regression models, the overall fit of the model will be assessed (e.g., Hosmer-Lemeshow Test). Receiver Operator Characteristic (ROC) curves will be generated and the area under the ROC curve will be examined.

The analysis of time to remission and time to response will be analyzed using standard survival analysis techniques. For discrete predictors variables (e.g., treatment), Kaplan-Meier curves will be generated and a log-rank statistic will be used to compare the distribution of the curves. For continuous predictor variables (e.g., age), Cox-Proportional Hazards models will be used to estimate the bivariate relationship. Cox-Proportional Hazards models will also be used to identify factors independently associated with the time to the event of interest. As was proposed for the logistic regression analyses, multiple regression models will be generated to determine if there is an independent effect of a variable of interest (e.g., treatment) after adjusting for possible confounding variables. The confounding variables will be identified through literature review or analysis of the baseline characteristics. In exploratory analyses, a stepwise or a backward stepping approach will be utilized to identify factors independently associated with the time-to-event outcome.

The proportionality assumption will be assessed for all Cox Proportional Hazards models. If the proportionality assumption is violated, alternate approaches will be investigated. This would include conducting a split-analysis of the data, where the time would be divided into intervals where the hazard functions are proportional within the time intervals.

Other approaches that may be utilized through the course of the study include ANOVA, ANCOVA and linear regression. For example, the analysis of the data for Specific Aim C-1 would utilize these approaches to assess the effect of treatment on function, as measured by the SF-12. For each of these approaches, the necessary distributional assumptions will be tested. If the distributional assumptions are not met, transformations will be sought. For example, log transformations of costs are frequently required to normalize residuals in standard regression models to obtain consistent estimates and valid significance tests of parameter coefficients to predictor variables. If suitable transformations cannot be identified, nonparametric approaches, such as a nonparametric ANOVA (Kruskal-Walis) and nonparametric regression techniques (e.g., Classification and Regression Trees (CART)) will be utilized.

A summary of the analytic approach for each of the study hypotheses is provided in the following table.

Table 9. Analytic Approaches for Specific Aims/Objectives

SpecificAim/ Objective	Analytic Approach
A-1	Survival analysis
A-2 (1)	Chi-square test, logistic regression
A-2 (2)	Chi-square test, logistic regression
A-2 (3)	Chi-square test, logistic regression
A-2 (4)	Chi-square test, logistic regression
B-1 (1)	Chi-square test
B-2 (1)	Rates, confidence intervals
B-2 (2)	Logistic regression
B-2 (3)	Chi-square test, logistic regression
B-2 (4)	Chi-square test, logistic regression
C-1 (1)	Chi-square test, logistic regression, ANOVA, linear regression
C-2 (1)	Rates, confidence intervals
C-2 (2)	Chi-square test, logistic regression
C-2 (3)	Chi-square test, logistic regression
C-2 (4)	Chi-square test, logistic regression
D-1 (1)	Chi-square test, logistic regression
D-1 (2)	Chi-square test, logistic regression, ANOVA, linear regression
D-1 (3)	Chi-square test, logistic regression, ANOVA, linear regression
D-2 (1)	Rates, confidence intervals
D-2 (2)	Chi-square test, logistic regression
D-2 (3)	Chi-square test, logistic regression
D-2 (4)	Chi-square test, logistic regression
E-1 (1)	Chi-square test, logistic regression
E-1 (2)	ANOVA, linear regression
E-2 (1)	Rates, confidence intervals
E-2 (2)	Chi-square test, logistic regression
E-2 (3)	Chi-square test, logistic regression
E-2 (4)	Chi-square test, logistic regression
F-1 (1)	t-test, random regression
F-1 (2)	t-test, random regression
F-2 (1)	t-test, random regression
F-2 (2)	t-test, random regression

Specific Data Analysis Approach

The primary goal of the proposed project is to assess treatment differences in Level 2 of the study. Because of this, the overall study is powered on the analysis of Level 2. It is important that the analysis approach for this primary specific aim be presented (Hypothesis B-2 (1)) in detail. The justification for such an analytic approach lies in the available sample size and the size of the effect. A description of these conditions is provided in the Power and Effect Size section.

A chi-square test will be used to make pair-wise comparisons among the treatment options within the medication switch subset, and among the medication augment subset. If no significant differences are detected within a subset, the data within the subset will be pooled. Otherwise the most effective treatment will be used at the next level of testing. For example, if the pair-wise comparisons of the medication switches showed that SER was the most effective of the switches, only those assigned to SER will be used in the comparison of medication switch to the other strategies (i.e., medication augment, psychotherapy switch, psychotherapy augment). If the pair-wise comparisons of the medication switches showed that no one medication switch was more effective than the others, then all the participants assigned to a medication switch will be used in the comparison of the medication switch to the other strategies (i.e., medication augment, psychotherapy switch, psychotherapy augment).

Next all pair-wise comparisons will be made between the medication switch, medication augment, psychotherapy switch, psychotherapy augment groups. As was indicated earlier, the medication switch and medication augment groups may be pooled depending on the results of earlier analyses designed to test for treatment differences within each of the strata.

Data Analysis Issues

1. Clinical Site Effects
In multi-center studies such as this, there is always the possibility of a site-effect. That is, the there is a possible site by treatment interaction that would indicate that the treatment effect is not generalizable. In STAR*D, in addition to the possible site-effect, there is also a clinical site, within Regional Center (RC)-effect. These effects will be assessed using the appropriate hierarchical models that will estimate and control for the possible correlation of subjects within a level.

2. Estimability of Treatment Effect Contrasts
The general format of each level in the tree of treatment strategies can be formalized as a two step process: first determine the participant's "acceptability profile," which defines a stratum and a set of admissible treatments for that stratum, and then randomize to one of the admissible treatments. For each such acceptability stratum, the outcomes of randomized participants provide unbiased estimates of the stratum-specific contrast of effects between pairs of treatments in the admissible set.

For example, if all participants in stratum 1 accept (and are indifferent to) each of the treatments A, B, and C, then the observed outcomes provide estimates of the relative effects of A versus B, B versus C, and A versus C. More concretely, the observed difference in recovery rates in participants assigned to A and B, respectively, estimates the comparative effect of A versus B, in stratum 1, which we can denote by [AvB;s=1]. This is because the groups of participants within this stratum who are assigned to each treatment are balanced with each other (up to chance) on all prognostic features, by the randomization. In this sense, the stratification by "acceptability" is no different than any other fixed and measured prognostic factor, such as site, sex, or previous history of treatment.

The clinician's judgment, as well as the participant's preferences, should be included in the acceptability stratification. For example, if the clinician knows that the participant is likely to be allergic to a medication (perhaps based on a history of reaction) or is not likely to comply with a particular medication because of a common side effect, or is not likely to be a suitable candidate for a particular form of medication switch (eliminating all medication switches), then that information can and should be used to define the acceptability stratum. In this way, the method is identical to the way that inclusion and exclusion criteria, together with informed consent, define the set of eligible-for-randomization participants in the typical two or three group treatment trial. The difference is that, in the usual trial, participants who are ineligible for any reason are not studied further. In this trial, we will attempt to find a randomized experiment that fits each participant. Of course, a participant whose acceptability stratum is defined by the empty set of admissible treatments, or a singleton, among the options under study, will not be suitable for this trial either, until that situation changes.

3. Pooling Across Strata

Suppose two strata (1, and 2) are defined by the acceptability profiles {A,B,C} and {A,B,D} respectively. Then we have unbiased estimates of the effects [AvB;s=1] and [AvB;s=2]. These can be combined ("pooled") to form an unbiased estimate of the effects of treatment A versus B in a population that includes a mixture of the two kinds of participants, those from stratum 1 and those from stratum 2 (for example, the average of the two effects weighted by the relative sizes of the strata in the original sample). That is, for the comparison of AvB, we can combine, or pool, the data from s=1 and s=2.

In this way, we can obtain a "broadly based" estimate of the A versus B effect in all participants who would entertain either A or B. It is clear that this is the only population for whom the effect of A vs. B is currently relevant, since participants who exclude either A or B do not have a well-defined effect of A vs. B, even in concept. For example, [CvD;s=1] is not well defined, and thus is not estimable.

Note further that the participants in stratum 1 who are randomized to C and those in stratum 2 who are randomized in D cannot be compared on outcome to estimate a C vs. D effect, since they are separated by stratification. However, if there is a stratum 3 with admissible treatment list{C,D,....} including C and D, then the contrast [CvD;s=3] is estimated without bias.

By pooling estimates across strata, each estimable contrast is estimated as efficiently as possible, given the a priori distribution of acceptabilities. Thus, a pair of options that is acceptable to many individuals, regardless of their other preferences, will be well estimated, while option pairs that seldom appear together in acceptability profiles will be poorly estimated (in the sense of experimental error and power to detect differences). If we want to test many options we must spend participants on them. In the simplest of all possible worlds, we would find that all treatment options are equally acceptable a priori, and this would all reduce to a K-arm randomized experiment, with K treatment options.

There is some structure to the treatment options; for example, A might be "augment CIT with BUS" while B might be "augment CIT with BUP," and C, D and E might be three switching

options, respectively SER, BUP, and VEN. Then it is possible to take averages of treatment contrasts to answer the question "is augmentation better than switching?" The structure described above helps to sharpen the question. For example, suppose the target of estimation is the relative success rate of the following two "strategies": (1) switch a participant, choosing the new treatment at random from among the participant's admissible switching options; (2) augment with a treatment chosen at random from the participant's admissible augmentation options. Then it is clear that this is only well defined for participants who are willing to accept all medication switch and medication augment strategies. Such participants belong to strata with at least one of A,B and at least one of C,D,E.

For example, if stratum 1 has admissible treatments {A,B,C,D,E} we obtain [AvC;s=1], [AvD;s=1], [AvE;s=1], [BvC;s=1], [BvD;s=1], [BvE;s=1]. These can be averaged to compare the effects of the "compound" strategies of augmentation versus switching in this stratum, which consists of participants who would find all medication switch and augment strategies acceptable.

In principle it is possible to estimate effects of more deterministic strategies. For example, suppose one entertained the strategy (1') switch to the first medication that the participant would find acceptable in the following list {SER, BUP, VEN}, and compared it to a comparably ordered strategy for augmentation (2') augment with BUS if acceptable, otherwise BUP. Then reweighting the stratum-specific contrasts can be employed to estimate the contrast of (1') and (2').

4. Statistical Outcome and Clinical Failure

In the above, we have assumed that some criterion for success or failure will be used to define the outcome for statistical purposes. However, the participant may also "fail" in the sense that the clinician believes that the participant should no longer be treated in the current manner, and it is necessary to declare failure and move the participant to the next level of treatment options. As long as these two concepts of failure are concordant, there is no difficulty. However, if a clinical failure occurs before the participant is judged a "statistical" failure (e.g., treatment stopped before adequate trial achieved), then the following problem arises. From the "intent-to-treat" point of view, the outcomes that occur after clinical failure but before statistical failure, should be counted toward the treatment under study in the current level. Thus, a participant randomized at level 2 to SER switching who fails clinically but not statistically, and is then randomized to VEN and dramatically improves, might be counted as a Level 2 SER success, by ITT. Dropping that participant from the analysis because of censorship at treatment change violates ITT. Defining that participant as a statistical failure removes the discordance, by defining it away, but may do violence to the principle underlying blinded outcomes assessment (the clinician's behavior is now the determinant of the outcome, statistically). At the same time, the ITT estimate may not be satisfying to the clinical participant of the research.

There is no "royal road" to a resolution of this problem, which is generic to all trials. However, it must be explicitly considered before the data analysis can commence. A more complex situation arises if the outcome of the participant is not dichotomized, but rather depends on the longitudinal course of symptoms, in a "repeated measures" paradigm. Then

the clinical failure before the fixed end-of-trial (say, 8 or 12 weeks) induces a missing data or non-compliance effect into the ITT method. However, the use of longitudinal data should make the contrasts more powerful, statistically.

5. Interim Analyses

The DSMB will monitor, among other things, recruitment at each of the clinical sites, protocol deviations, adverse events, descriptive characteristics of the population at baseline by treatment group and effectiveness of the treatments at the various levels of the study. Official interim monitoring results of the study will be presented for review. To avoid potential bias, participating practitioners will not review the results of the interim analyses.

A number of different group sequential analysis approaches are available for interim analyses of the effectiveness of treatment. We propose an approach that adjusts the type I error rate in a non-uniform fashion, so that the majority of the type I error is conserved for the final treatment comparison after recruitment has been completed. Thus, large difference would need to exist early in the trial to find a significant difference during the monitoring phase.

Two possible approaches would be the O'Brien and Flemming approach, and the Lan and DeMets approach. With the O'Brien and Flemming approach, the number of participants having entered the study and completed the necessary follow-up would be fixed and analyses would be conducted after the these thresholds are met (e.g., after every 200 participants complete follow-up). This is highly dependent on the recruitment rates and it is not known if these recruitment and follow-up milestones would correspond with the dates of the DSMB meetings. Because this, we propose to use the Lan and DeMets approach which spends the type I error as a function of time, and the number of interim analyses does not need to be specified *a priori.*

Power and Effect Size

Level 2

The sample size necessary for Level 1 was determined by the need for an adequate number of participants entering Level 2. A 50% failure rate is expected for Level 1 and it is estimated that 2,000 participants are needed for Level 2, therefore 4,000 participants will be needed at Level 1.

To determine the detectable effect size we must first lay out the general analytic approach and the rationale for that approach. There are seven treatment options altogether in the Level 2 analysis. If we simply compared all pairs of treatments, we would have to correct for 7*6/2 comparisons, which would be rather severe. Instead, we propose a "step up" procedure. Taking advantage of the fact that there are two strategies (medication switch and medication augment) where we can organize outcomes ("recovery" proportions): p 1, p 2, p 3 for the three medication switches and p 4 and p 5 for the two medication augments. In addition, there are the recovery proportions from psychotherapy switch (p 6) and psychotherapy augment (p 7).

First, test the three null hypotheses corresponding to equal within-strategy success proportions: (p 1=p 2, p 2=p 3, p 4=p 5), at level alpha/3 (two sided). Call this "Test 1."

Within each strategy, if any of them is rejected, take the subset winner and compare them, at level alpha/6 (there are 4*3/2=6 pair wise comparisons). (These may involve some pooling of contrasts with equivalent options within an approach). Call this "Test 2". If there are no within-strategy successes, compare the 4 strategy groups pooling within the medication switch and medication augment strategies at level alpha/6 (there are 6 =4*3/2 pairs). Call this "Test 3". Then the only way we declare an overall winning option (or winning approach) is if we get a rejection of a Test 2 or Test 3 hypothesis.

It is possible that the within strategy comparison would yield a success for one strategy (e.g., medication switch) and no successes in another strategy (e.g., medication augmentation). In this situation, both Test 2 and Test 3 would be utilized. The data within the medication augmentation and the winner of the medication switch strategy would be used for all the pairwise comparisons.

Under the full null hypothesis, p 1=p 2,..., =p 7 the overall type I error rate (probability of rejecting at least one null in Test 2 or Test 3) = Prob(Test1 rejects) Prob(Test 2 rejects | Test 1 rejects) + Prob(Test 1 does not reject) Prob(Test 3 rejects | Test 1 does not reject). But Prob(Test 1 rejects) < alpha, under the null, by Bonferroni, and Prob(Test3 rejects | Test 1 does not reject) < alpha, under the null, by Bonferroni, and the orthogonality of the within and between approach contrasts, so the overall Type I error rate < alpha*Probability(Test 2 rejects | Test 1 rejects) +alpha < (1+Probability(Test 2 rejects | Test 1 rejects))* alpha <2*alpha.

Thus, if alpha = 0.025, the overall rate is less than 0.05. This is a bit conservative, since it involves the nonsharp inequalities Prob(Test 1 does not reject) < 1 and Prob(Test 2 rejects | Test 1 rejects) < 1. But Prob(Test 1 does not reject) > 1-alpha (by Bonferroni), and Prob(Test 2 rejects | Test 1 rejects) may be large — it is not independent.

The most stringent contrasts involve Test 1, since there the tests are made at .025/3 = .0083, and they are the "smallest n" contrasts (all the others would involve at least some pooling). To get an idea of the inflation in N this entails, compared to the usual .05 level, at 80% power, we will require about 1.5 times the N. So we can think about powering tests as usual, and then multiplying the N by 1.5, to get the right power for the conservative contrasts. By contrast, if we use a 21-fold Bonferroni for the 7*6/2 basic pair wise contrasts, we will need about 1.9 times the N.

It should be borne in mind that these are the "effective sample sizes" for a contrast, taking into account the fact that several strata contribute to a contrast. Based on the acceptability data provided in Figure 6, we can estimate that there will be 287 participants in each of the medication switch cells, 354 participants in each of the medication augment cells, 204 participants in the CT switch cell and 224 participants in the CT augment cell.

These sample sizes can then be used to identify the effect size that can be detected in Test 1 from above. Using a chi-square test for association, and assuming 80% power, a type I error rate of .025/3=.0083, a two-sided alternative hypothesis, a sample size of 287 per cell in the medication switch stratum and a sample size of 354 in the medication augment cells, an effect size of .205 can be detected for the medication switch comparisons and an effect size of .185 can be detected for the medication augment comparison.

The next step will depend on the results of Test 1. If differences are found in Test 1 within either the medication switch or medication augment strata, then only those participants that received the more advantageous treatment within that stratum will be used in the following analyses, if not the results will be pooled within the stratum. Therefore, the sample sizes for the next set of tests will be based on the results of Test 1 and the data from Figure 6. The table below presents the sample sizes for each comparison in Tests 2 and 3, depending on the results of Test 1 (if significant differences are not detected in Test 1 within a stratum the results sample sizes are pooled, otherwise the sample size of the most effective treatment is used).

The numbers below were obtained from utilizing the estimated sample sizes from Figure 6. The number of participants expected for each acceptability stratum was divided by the number of treatment options available within the stratum. These were then summed across the strata with the same comparisons, either for a given treatment (Test 2) or for a given treatment strategy (Test 3).

Table 10. Effect Size by Comparisons at Level 2

Comparison Groups		Test 2			Test 3		
Group 1	Group 2	N_1	N_2	Effect Size	N_1	N_2	Min Effect Size
Med Switch	Med Augment	132	132	.228	396	264	.161
Med Switch	Psychotherapy Switch	93	93	.272	279	93	.278
Med Switch	Psychotherapy Augment	31	31	.471	93	31	.471
Med Augment	Psychotherapy Switch	112	112	.248	224	112	.248
Med Augment	Psychotherapy Augment	112	112	.248	224	112	.248
Psychotherapy Switch	Psychotherapy Augment	141	141	.221	141	141	.221

From the above table, we can see that if a more effective treatment is identified in the medication switch stratum and in the medication augment stratum, the total sample size will be 264 (132 in each group), and the effect size that can be detected with this sample size, using the chi-square test, and assuming 80% power, a two-sided alternative hypothesis and a type I error rate of .0042 (=.025/6) we will be .228. However, if a more effective treatment is not detected in either stratum, the samples will be pooled within each stratum yielding a sample size of 396 in the medication switch stratum and 264 in the medication augment

stratum. If we assume a balanced design and conservatively use the smaller sample size of 264 for both arms (total sample size of 528), an effect size of .161 can be detected using the chi-square test, assuming 80% power, a two-sided alternative hypothesis and a type I error rate of .0042 (=.025/6).

Table 11 presents the size of the difference that can be detected between the response rates. The five characteristics of a study that can affect the size of the difference of the responses rate to be detected are the error rates, both type I and type II, the direction of the alternative hypothesis (one sided versus two sided), the sample size, and the response rate in one of the groups. In this instance, the type I error rate is set to .0042, the type II error rate is set to .20 and the direction of the alternative hypothesis is two-sided. The sample sizes for various comparisons have been fixed, so Table 8 presents the difference to be detected for various levels of response in one of treatment groups. For example, for the comparison of medication switch to medication augment when a "best" medication switch is identified and a "best" medication augment is identified has a sample size of 132 in each cell. For the comparison of medication switch to medication augment, have sufficient power to detect a change in the response rate of +/- .226, if the response rate of one of the treatments is .4.

Table 11. Difference in Response Rate to be Detected at Level 2

N/cell	P	)	N/cell	P	)	N/cell	P	)
31	.1	.407	112	.1	.190	141	.1	.166
31	.2	.467	112	.2	.226	141	.2	.200
31	.3	.453	112	.3	.242	141	.3	.215
31	.4	.437	112	.4	.245	141	.4	.218
31	.5	.404	112	.5	.238	141	.5	.214
93	.1	.213	132	.1	.173	264	.1	.116
93	.2	.250	132	.2	.207	264	.2	.142
93	.3	.266	132	.3	.222	264	.3	.135
93	.4	.268	132	.4	.226	264	.4	.160
93	.5	.258	132	.5	.220	264	.5	.158

N/cell=Number of participants in each of the two cells
P=Reponse rate in one group
)=Size of difference that can be detected

Level 2
Assuming a sample size of 170 (85 per group), a type I error rate of .05, a type II error rate of .20, and a two-sided alternative hypothesis, an effect size of .215 can be detected among the response rates of the two treatment groups.

Level 3

An approach similar to the approach of the analysis for Level 2 will be taken for the analysis of the Level 3 data. In this instance the number of acceptability strata and the number of comparisons are smaller. A total of 698 participants are expected to enter Level 3. If we assume 25% will choose a switch strategy, 25% will choose an augment strategy and 50% will be willing to accept all possible treatments. There will be only one teat (e.g., Test 1) in each of the acceptability stratum to determine if there is a more effective treatment in the switch preference and in the augment preference. A Type I error rate f .025/2 will be used for each of these tests. The next step will be to compare all four possible treatments using the Test 2 and Test 3 approach as described above. The table below presents the sample sizes for each comparison in Tests 2 and 3, depending on the results of Test 1 (if significant differences are not detected in Test 1 within a stratum the results sample sizes are pooled, otherwise the sample size of the most effective treatment is used). A Type I error rate of .025 will be used for Test 2 or Test 3. This error rate is larger than the error rate used for Tests 2 and 3 in Level 2 due to the fact that there are fewer acceptability strata, and therefore, fewer comparisons.

Table 12. Effect Size by Comparisons at Level 3

Comparison Groups		Test 2			Test 3		
Group 1	Group 2	N_1	N_2	Effect Size	N_1	N_2	Min Effect Size
Switch	Augment	87	87	.234	174	174	.146

Level 4

Assuming a sample size of 292 (146 per group), a type I error rate of .05, a type II error rate of .20, and a two-sided alternative hypothesis, an effect size of .164 can be detected among the response rates of the two treatment groups.

Generally, an effect size of 0.2 is considered small, while 0.5 is considered moderate. Therefore, at Levels 2, 2A, 3 and 4, , the study has the necessary power to detect between small and moderate effects.

V. Data Coordination

Data Management System

An integration of fax and optical character recognition (OCR) technology based at the EDC comprises the main core of the data collection system. Blank data collection forms are downloaded from the STAR*D website and printed. The forms are then completed during participant interviews. Upon completion, forms are faxed by the CRC or the ROA via a toll-free number to a fax server located in the EDC. The form image is processed at the EDC to extract data fields for transfer to the central database. EDC Data management staff perform validation procedures to ensure correct interpretation of all scanned data before loading into the database.

Additional data collection is performed through interactive voice recognition (IVR) software designed and maintained by HealthCare Technologies Systems (HTS) in Madison, Wisconsin. IVR data are sent to the EDC via weekly standard file transfer protocol (FTP) procedures, and are added to the database upon receipt. In addition, standard faxes automatically produced by the IVR system to notify study personnel of participant enrollment and randomization are sent to the EDC's fax server for processing .

The central database management system is maintained through the SAS system, and is comprised of several databases. One DYNAMIC database contains a table for every type of form or IVR file received. This dynamic database is constantly changing during data accrual and the editing process.

If a form must be re-faxed to make a correction in response to edit checks, its data replaces the original in the dynamic database, and the original data are stored in a parallel AUDITS database to maintain an accurate trail of changes made to data during the editing process.

On a periodic basis, data that have passed all quality control procedures are extracted to a static ANALYSIS database which contains tables for all of the assessment measures as well as tables of general demographic and clinical information. This unchanging database will be used for interim analyses while additional data accrue in the dynamic database.

Study Management and Communications

Data management staff from the EDC will participate in the regularly scheduled CRC conference calls. The purpose of these calls will be to discuss participant recruitment, protocol implementation, and data collection and transfer. In our experience, these calls have helped to foster an interdependent, collaborative relationship among the participants. Communication during the recruitment phase of the trial will be particularly important. Many questions are expected as the CRCs become accustomed to implementing the protocol.

The core of the information management system and communications will be a web site. The web site will be designed with features to promote the accurate and efficient operation of the study and will serve as a central location for retrieving and disseminating information. *News*

bulletins, operations memos, calendars, personnel directories and frequently asked questions are example features that are accessible via the STAR*D web site. Additionally, dynamic reports including data integrity and compliance reports are available for review. A shared documents section will provide a restricted area for adding and retrieving shared manuscripts, operations memos, training documentation, data collection forms, and various other reports.

Web site features are categorized into two areas based on the level of security required for access. These areas are public and private. The public Web site options are open to the general Web visitor. Public options include a home page with a project description, a bibliography page, and current news regarding the project. Private area features include the personnel directory with options such as individual and group electronic mail, telephone/voice mail, and U.S. mail information. A shared documents section provides a restricted area for operations memos, training documentation, manual of operations, news bulletins, data collection forms, and reports. This area can also be used for sharing manuscripts and for submitting help requests or queries to the EDC and the NCC. Additionally, dynamic utilities, such as a calendar in a month-at-a-glance format, provide information regarding scheduled events. Calendars can also be placed in the public area if desired to publicize events of interest to the general Internet user community. Password authentication and Internet Protocol address (IP) verification via the Web server will be necessary to access the private area.

Quality Control

Data that are loaded into the dynamic database are subjected to stringent quality control procedures that identify standard missing, range, and dependency errors, as well as performing logical checks to ensure correct chronology and expected relationships between related items. Records that contain inconsistent data loaded back onto blank forms in Adobe PDF format, with the fields in question marked and annotated. The PDF forms are e-mailed to the person who initially completed them, and corrections may be made directly on the electronic forms. A "reply" e-mail directs the corrected form images to the EDC's fax server where, flagged as updates, they join other images to be processed.

Upon receipt of an enrollment notification fax from the IVR system, the data management system automatically creates an individualized participant schedule to provide a framework by which timely data collection according to the study protocol may be ensured. The schedule is sent to the CRC and to HTS to confirm the timing of expected IVR calls. Subsequent forms received for that participant are compared to the schedule to determine compliance with data collection windows, and weekly reports are sent to the NCC and the CRC to identify overdue forms or IVR calls.

When a participant changes treatment levels or moves into followup, the participant schedule is modified to reflect the "restarted" clock, and is re-distributed to the CRC and to HTS.

Training

All ROAs and CRCs will be trained on the study protocol, methods of data collection, and data management. Training on the study protocol will cover the screening procedures, eligibility and exclusion criteria, data items to be collected, and the procedures for completing the data collection forms. Training on data management topics will include Web site access, data entry and management, correcting data, coding missing data, and running reports.

Certification will be required for all CRCs, and re-certification will be done annually through the study to ensure up-to-date knowledge of protocol additions or changes over the course of the study. Throughout the course of the project retraining and training for new personnel will be available. The tests used for certification will be administered at the via the website, if necessary, and will be repeated at regular intervals for re-certification.

Standard Operating Procedures

The NCC and the EDC will be responsible for the development of study materials, including the data collection forms and a Manual of Operations. They will develop, evaluate, revise, and produce standardized paper forms for collecting all data needed on study participants including eligibility, demographic and other baseline data, sequential clinical assessments, dosing information, side effects, and outcome measures. The EDC will ensure that the data collection forms are clear, concise, and easy to use. Strategies such as check boxes and different font and typefaces for clarity will be employed. The EDC will work with the investigators to ensure that all data required to explore the study hypotheses are collected and in the appropriate format for data analysis.

The Manual of Operations (MOP) will include an overall description of the project, the study protocol, data collection procedures, data management documentation and instruction, and detailed descriptions of all other procedures required to successfully implement the study. The MOP is a reference and training manual for the data collection personnel. The process of data collection will be documented in detail, including the provision of clear definitions and instructions for completing the data forms. Each study form will be listed, the data collection schedule will be provided, and copies of each form will be included in this portion of the MOP. Detailed instructions for completing each item of each form will be given. All study variables will be defined to assure consistency across data collectors. These definitions will be compiled alphabetically in a data dictionary and included in the MOP. All coded responses to variables will be compiled alphabetically in a codebook to be included in the MOP.

The MOPs will help to ensure smooth operation throughout the course of the study and provide some protection against potential disruptions due to personnel.

Reports

During the course of the study several types of technical and statistical reports will be prepared. Reports will be provided via the WWW site to study investigators on:
- recruitment activities at each clinical site and Regional Center
- participant retention

- forms received for each participant at each time-point
- schedule of upcoming interviews
- quality of data received by the EDC by the clinical sites (e.g., % missing data)
- data edits (e.g., data inconsistencies).

Clinical Site Monitoring

The quality of the data at the clinical sites will be monitored in several ways. First, data edit routines will be written to monitor the accuracy of the data. These routines will identify and inconsistencies within and across forms. These reports will be generated on a routine basis and submitted to the clinical sites for correction. The clinical sites will then be monitored for a timely repose to these reports. Additional reports will be generated to monitor the clinical sites. These will include reports generated to identify the data that have been received and indicate data that are currently missing from the database for a given participant.

If problem clinical sites are identified, these sites will be visited to assure that data are being collected in an accurate fashion. During the clinical site visits, the sites will be monitored to assure that they are collecting data in accordance with the protocol. To assure this, mock interview will be scheduled or the tapes of previous interviews will be reviewed. In addition, the data will be audited. Any inconsistencies will be investigated and corrected.

Security

The EDC enforces a comprehensive security policy. The EDC network defenses include IP restriction, basic or digest authentication, and elimination of anonymous FTP to protect the data system and database from common network intrusions. A firewall will provide an additional barrier between the Internet community and the EDC network and the study database. The firewall will restrict, monitor, and log Internet traffic traveling in and out of the EDC, and will alert system administrators of potential threats to security.

Regularly scheduled backups and archives at the EDC will protect central and local information from hard disk failures. Permanent archives of critical project files are created using various media (e.g., DAT, CD-ROM, etc.) and are stored in a secured off-site facility to prevent data loss due to catastrophic events. Routine virus detection and virus shields will be enforced for all EDC computers. All critical information regarding database transactions will be logged and stored in journal files. In the event of inappropriate use of the project database, a previous database state may be restored from backup media or journal files. All servers used for this project will be connected to uninterrupted power supplies to protect equipment and data from damage resulting from electrical surges and outages. A secured, raised-floor computer room in an area with a burglar alarm will house all project server equipment.

VI. Human Subjects

Male and female depressed outpatients aged 18-75, for whom antidepressant treatment is deemed appropriate by their treating clinician, with major depressive disorder (MDD) will be enrolled. We expect to enroll 4,000 such participants in the treatment protocol in order to generate 2,000 protocol-treated participants who have not benefited sufficiently from citalopram (CIT) — the Level 1 treatment — such that further treatment is called for by their clinician and this protocol.

The major subpopulations excluded are those <18 years of age and those over age 75 at study entry. The younger group is excluded because of the paucity of studies documenting antidepressant medications' safety, efficacy, and appropriate dosing in this group, together with the need for specialized providers of cognitive therapy and for specialized pediatric expertise and instruments among younger participants.

We excluded those over age 75 because of the high likelihood of encountering concomitant general medical conditions (GMCs) or concomitant medications that would make some, but not other, treatment options at Level 2, 2A, 3, and 4 strongly preferred for safety reasons. Secondly, this older group of participants differs from younger adults with regard to the rate of dose increases, preferred starting doses, speed of response, and medication metabolism, making it likely that a different response pattern may be encountered, yet insufficient sample size will be obtained to adequately study these differences.

Pregnant women and women who are breast-feeding are excluded because of the potential unknown risk of medications to the fetus and infant respectively.

Figures 8, 9, and 10 show the steps involved for participants at screening (Figure 8), protocol clinic visits (Figure 9) and for all participants in follow-up (Figure 10).
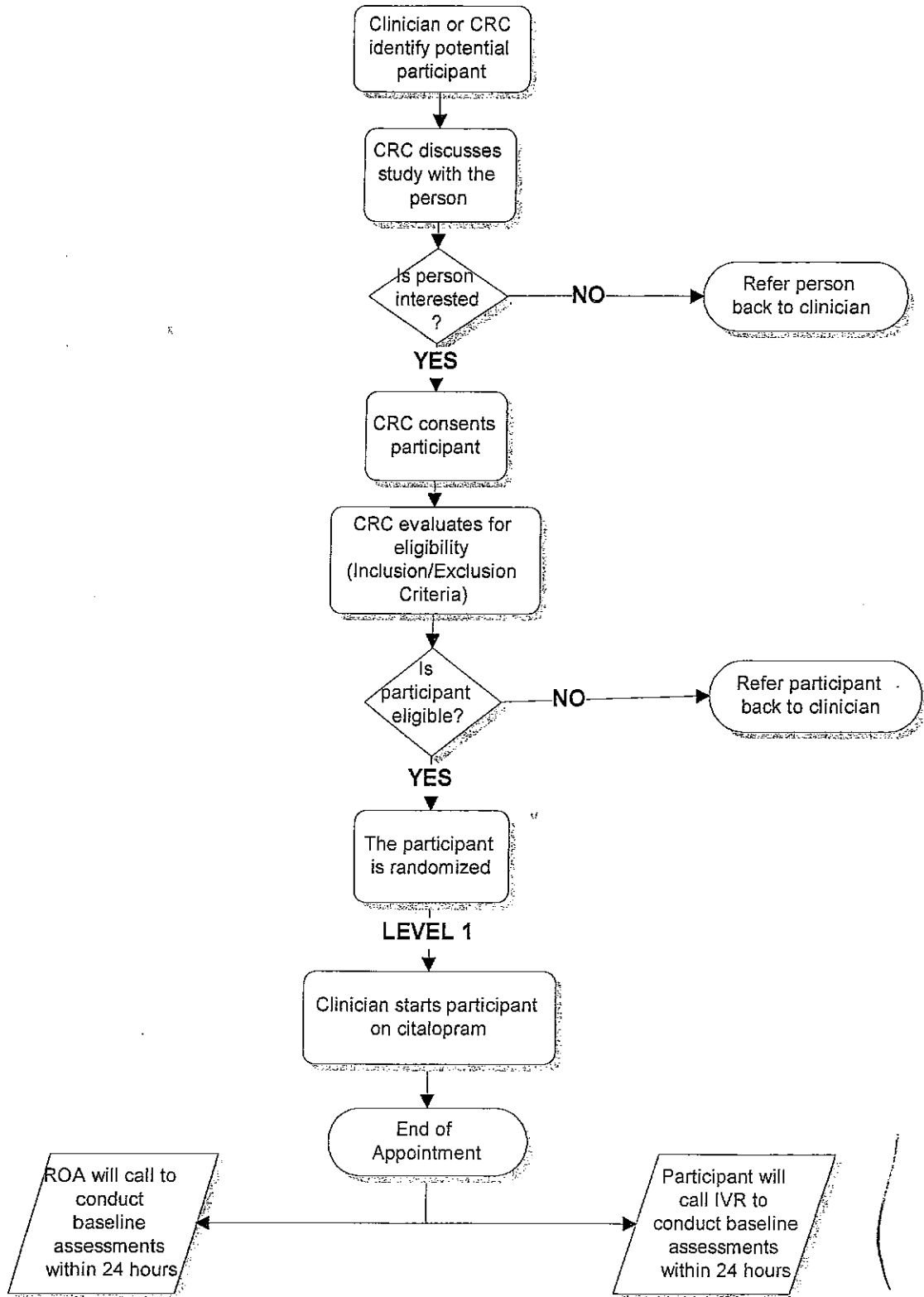
Figure 8: Screening/Baseline Visits

```
                    ┌─────────────────┐
                    │ Clinician or CRC│
                    │ identify potential│
                    │   participant   │
                    └────────┬────────┘
                             │
                             ▼
                    ┌─────────────────┐
                    │  CRC discusses  │
                    │  study with the │
                    │     person      │
                    └────────┬────────┘
                             │
                             ▼
                        ◇─────────◇
                       ╱ Is person  ╲           ┌──────────────────┐
                      ◇  interested  ◇──NO────▶ │  Refer person    │
                       ╲     ?      ╱           │  back to clinician│
                        ◇─────────◇             └──────────────────┘
                            │
                           YES
                            ▼
                    ┌─────────────────┐
                    │  CRC consents   │
                    │   participant   │
                    └────────┬────────┘
                             │
                             ▼
                    ┌─────────────────┐
                    │ CRC evaluates for│
                    │   eligibility   │
                    │(Inclusion/Exclusion│
                    │    Criteria)    │
                    └────────┬────────┘
                             │
                             ▼
                        ◇─────────◇
                       ╱    Is     ╲            ┌──────────────────┐
                      ◇ participant  ◇──NO────▶ │ Refer participant│
                       ╲  eligible? ╱           │ back to clinician│
                        ◇─────────◇             └──────────────────┘
                            │
                           YES
                            ▼
                    ┌─────────────────┐
                    │ The participant │
                    │  is randomized  │
                    └────────┬────────┘
                          LEVEL 1
                             │
                             ▼
                    ┌─────────────────┐
                    │Clinician starts participant│
                    │   on citalopram │
                    └────────┬────────┘
                             │
                             ▼
                    ┌─────────────────┐
                    │     End of      │
                    │   Appointment   │
                    └────────┬────────┘
```

ROA will call to conduct baseline assessments within 24 hours

Participant will call IVR to conduct baseline assessments within 24 hours

Figure 9: Clinic Visits at Each Level

Participant has clinic visits at weeks 2,4,6,9,12

CRC meets participant and completes ratings and patient education

CRC enters findings on CRF

Participant sees clinician

Is it week 6? — YES → Participant will call IVR to conduct mid-level assessments

ROA will call to conduct HAM-D17, IDS-C30, and IPAQ ← YES — Is it exit of level? — YES → Participant will call IVR to conduct Exit of Level assessments

NO

Is depression remitted — NO → Go to next level

YES

Go to Follow-up

Figure 10: Post-Level Follow-up
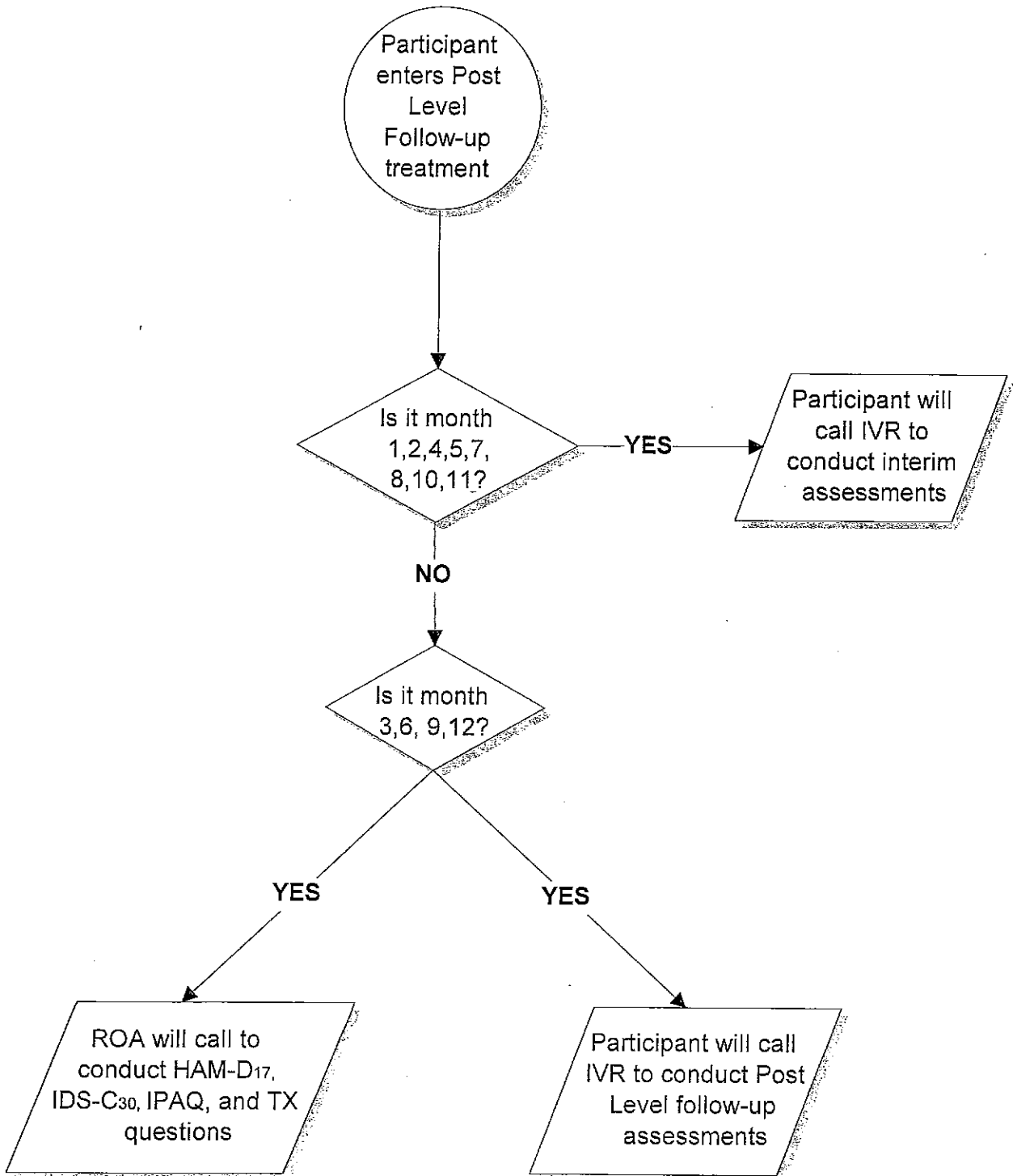
List of Acronyms

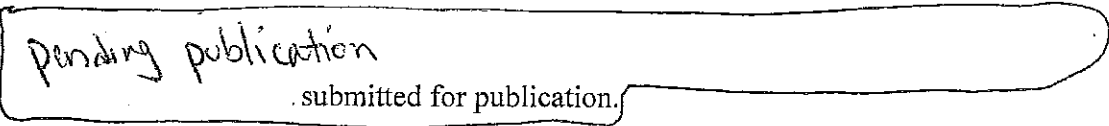|		
BUP	—	BupropionSR
BUS	—	Buspirone
CGI-I	—	Clinical Global Inventory – Improvement
CIRS	—	Cumulative Illness Rating Scale
CIT	—	Citalopram
CRC		Clinical Research Coordinator
CRF	—	Clinical Record Form
CT	—	Cognitive Therapy
EDC	—	Data Center
FISER	—	Frequency and Intensity of Side Effects Ratings (2-item)
GMC	—	General Medical Condition
GRSEB	—	Global Rating of Side Effects Burden (1-item)
$HAM\text{-}D_{17}$	—	Hamilton Rating Scale for Depression (17-item)
$IDS\text{-}C_{30}$	—	Inventory of Depressive Symptomatology – Clinician-Rated (30-item)
IPAQ	—	Income and Public Assistance Questionnaire
IVR	—	Interactive Voice Response
Li	—	Lithium
MAOI	—	Monoamine Oxidase Inhibitor
MDD	—	Major Depressive Disorder
MIRT	—	Mirtazapine
NCC	—	National Coordinating Center
NTP	—	Nortriptyline
PRISE	—	Patient Rated Inventory of Side Effects
PSI	—	Patient Satisfaction Inventory
$QIDS\text{-}C_{16}$	—	Quick Inventory of Depressive Symptomatology – Clinician-Rated (16-item)
$QIDS\text{-}SR_{16}$	—	Quick Inventory of Depressive Symptomatology – Self Report (16-item)
Q-LES-Q	—	Quality of Life Enjoyment and Satisfaction Questionnaire
RC	—	Regional Center
RCT	—	Randomized Controlled Trial
SER	—	Sertraline
SF-12	—	Short-Form Health Survey (12-item)
SSRI	—	Selective Serotonin Reuptake Inhibitor
STAR*D	—	Sequenced Treatment Alternatives to Relieve Depression
TCA	—	Tricyclic Antidepressant
TCP	—	Tranylcypromine
THY	—	Thyroid Hormone
TX	—	Treatment Questions
UAC-IVR	—	Utilization and Cost Methodology Interactive Voice Response
UAC-PQ	—	Modified Utilization and Cost Patient Questionnaire
VEN	—	VenlafaxineXR
WPAI	—	Work Productivity and Activity Impairment Questionnaire (6-item)
WSAS	—	Work and Social Adjustment Scale (5-item)

Bibliography

1. Akiskal HS. Subaffective disorders: dysthymic, cyclothymic, and bipolar II disorders in the "borderline" realm. Psychiatr Clin North Am 4:25-46, 1981.

2. Amemiya T. Advanced Econometrics. Cambridge: Harvard U Press, 1985.

3. American Psychiatric Association. Practice guideline for major depressive disorder in adults. Am J Psychiatry 150(Suppl 4):1-26, 1993.

4. Ancill RJ, Rogers D, Carr AC. Comparison of computerized self-rating scales for depression with conventional observer ratings. Acta Psychiatr Scand 71:315-317, 1985.

5. Anderson FA, Wheeler HB, Goldberg FJ, Hosmer DW, Forcier A, Patwardhan NA. Changing clinical practice: Prospective study of the impact of continuing medical education and quality assurance programs on use of prophylaxis for venous thromboembolism. Arch Intern Med 154:669-677, 1994.

6. Austin MPV, Souza FGM, Goodwin GM. Lithium augmentation in antidepressant-resistant patients: A quantitative analysis. Br J Psychiatry 159:510-514, 1991.

7. Basco MR, Rush AJ. Compliance with pharmacotherapy in mood disorders. Psychiatr Ann 25:269-270,276,278-279, 1995.

8. Beasley CM Jr, Sayler ME, Cunningham GE, Weiss AM Masica DN. Fluoxetine in tricyclic refractory major depressive disorder. J Affect Disord 20:193-200, 1990.

9. Bentkover JD, Covello VT, Mumpower J, Spicker SF. Benefits Assessment: The State of the Art. Boston (MA): D. Reidel Publishing Company, 1986.

10. Berndt ER, Finkelstein SN, Greenberg PE, Howland RH, Keith A, Rush AJ, Russell J, Keller MB. Workplace performance effects from chronic depression and its treatment. J Health Econ 17:511-535, 1998.

11. Bielski RJ, Lydiard RB. Therapeutic trial participants: Where do we find them and what does it cost? Psychopharmacol Bull 33:75-78, 1997.

12. *pending publication* in press.

13. Bodkin JA, Lasser RA, Wines JD Jr, Gardner DM, Baldessarini RJ. Combining serotonin reuptake inhibitors and bupropion in partial responders to antidepressant monotherapy. J Clin Psychiatry 58:137-145, 1997.

14. Borus JF, Olendzki MC. The offset effect of mental health treatment on ambulatory medical care utilization and charges. Arch Gen Psychiatry 42:573-580, 1985.

15. Bouwer C, Stein DJ. Buspirone is an effective augmenting agent of serotonin selective re-uptake inhibitors in severe treatment-refractory depression. S Afr Med J 87(Suppl 4): 534-537, 1997.

16. Brennan PJ, Abrutyn E. Developing policies and guidelines. Infect Control Hosp Epidemiol 16:512-517, 1995.

17. Bridges PK. ". . . and a small dose of an antidepressant might help." Br J Psychiatry 142:626-628, 1983.

18. Brugha TS, Bebbington PE, MacCarthy B, Sturt E, Wykes T. Antidepressants may not assist recovery in practice. A naturalistic prospective study. Acta Psychiatr Scand 86, 5-11, 1992.

19. Bryk AS, Raudenbush SW, Congdon RT Jr. Hierarchical Linear and Nonlinear Modeling, with HLM/2l and HLM/3L Programs, ver. 4. Chicago (IL): Scientific Software International, Inc., 1996.

20. Bryk AS, Raudenbush SW. Hierarchical Linear Models. Newbury Park (CA): Sage Publications, 1992.

21. Byrne S, Rothschild AJ. Psychiatrists' responses to failure of maintenance therapy with antidepressants. Psychiatr Serv 48:835-837, 1997.

22. Carr AC, Ancill RJ, Ghosh A, Margo A. Direct assessment of depression by microcomputer. Acta Psychiatr Scand 64:415-422, 1981.

23. Charney DS. Monoamine dysfunction and the pathophysiology and treatment of depression. J Clin Psychiatry 59(Suppl. 14):11-14, 1998.

24. Chi EM, Reinsel GC. Models of longitudinal data with random effects and AR(1) errors. J Am Stat Assoc 84:452-459, 1989.

25. Clark RE, Ricketts SK, McHugo GJ. Measuring hospital use without claims: A comparison of patient and provider reports. Health Serv Res 31:153-169, 1996.

26. Cooper BS, Rice DP. The economic cost of illness revisited. Soc Secur Bull 39:21-36, 1976.

27. Cooper-Patrick L, Powe NR, Jenckes MV, Gonzales JJ, Levine DM, Ford DE. Identification of patient attitudes and preferences regarding treatment of depression. J Gen Intern Med 12:431-438, 1997.

28. Crismon ML, Trivedi MH, Pigott TA, Rush AJ, Hirschfeld RMA, Kahn DA, DeBattista C, Nelson JC, Nierenberg AA, Sackeim HA, Thase ME, Texas Consensus Conference Panel: The Texas Medication Algorithm Project. Report of the Texas Consensus Conference Panel on medication treatment of major depressive disorder. J Clin Psychiatry 60:142-156, 1999.

29. Cummings NA. Prolonged (ideal) versus short-term (realistic) psychotherapy. Prof Psychol 8:491-501, 1977.

30. Danton WG, Altrocchi J, Antonuccio D, Basta R. Nondrug treatment of anxiety. Am Fam Physician 10:161-166, 1994.

31. Davis DA, Thomson MA, Oxman AD, Haynes RB. Changing physician performance. A systematic review of the effect of continuing medical education strategies. JAMA 274:700-705, 1995.

32. Depression Guideline Panel. Clinical Practice Guideline. Number 5. Depression in Primary Care: Volume 1. Detection and Diagnosis. Rockville, MD: US Dept. of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research. AHCPR Publication No. 93-0550, 1993.

33. Depression Guideline Panel. Clinical Practice Guideline. Number 5. Depression in Primary Care: Volume 2. Treatment of Major Depression. Rockville, MD: US Dept. of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research. AHCPR Publication No. 93-0551, 1993.

34. Donabedian A. Benefits in Medical Care Program. Cambridge (MA): Harvard University Press, 1976.

35. Dornseif BE, Dunlop SR, Potvin JH, Wernicke JF. Effect of dose escalation after low-dose fluoxetine therapy. Psychopharmacol Bull 25:71-79, 1989.

36. Dresser MVB, Feingold L, Rosenkranz SL, Coltin KL. Clinical quality measurement: Comparing chart review and automated methodologies. Med Care 35:539-552, 1997.

37. Drummond MF, O'Brien BJ, Stoddard GL, Torrance GW. Methods for the Economic Evaluation of Health Care Programmes. Oxford (England): Oxford University Press, 1997.

38. Duan N, Manning WG, Morris CN, Newhouse JP. A comparison of alternative models for the demand for medical care. J Bus Econ Stat 1:115-126, 1983.

39. Duehrssen A. Katamnestische ergebnisse bei 1004 patienten nach analytischer psychotherapie. Zeitschrift fur Psychosomatische Medizin 8, 1962

40. Eddy MH. Clinical decision making: from theory to practice. Applying cost-effectiveness analysis. JAMA 268:2575-2582, 1992.

41. Eisenberg L. Treating depression and anxiety in primary care. N Eng J Med 326:1080-1083, 1992.

42. Eisenthal S, Emery R, Lazare A, Udin H. "Adherence" and the negotiated approach to patienthood. Arch Gen Psychiatry 36:393-398, 1979.

43. Elixhauser A, Halpern M, Schmier J, Luce BR. Health care CBA and CEA from 1991 to 1996: An updated bibliography. Med Care 36:MS1-MS9, 1998.

44. Elkin I, Shea MT, Watkins JT, Imber SD, Sotsky SM, Collins JF, Glass DR, Pilkonis PA, Leber WR, Docherty JP, Fiester SJ, Parloff MB. National Institute of Mental Health Treatment of Depression Collaborative Research Program. General effectiveness of treatments. Arch Gen Psychiatry 46:971-982, 1989.

45. Endicott J, Nee J, Harrison W, Blumenthal R. Quality of Life Enjoyment and Satisfaction Questionnaire (Q-LES-Q): A new measure. Psychopharmacol Bull 29:321-326, 1993.

46. Evans CE, Haynes RB, Birkett NJ, Gilbert JR, Taylor DW, Sackett DL, Johnston ME, Hewson SA. Does a mailed continuing education program improve physician performance? JAMA 255:501-504, 1986.

47. Ezra KR. An analysis of treatment preferences in patients of a late life depression clinic. Dissertation Abstracts International: Section B: The Sciences and Engineering 57(8-B):5324, 1997.

48. Fauman MA. Tricyclic antidepressant prescription by general hospital physicians. Am J Psychiatry 237:490-491, 1980.

49. Fava GA, Rafanelli C, Grandi S, Conti S, Belluardo P. Prevention of recurrent depression with cognitive therapy: Preliminary findings. Arch Gen Psychiatry 55:816-820, 1998.

50. Fava GA, Savron G, Grandi S, Rafanelli C: Cognitive-behavioral management of drug-resistant major depressive disorder. J Clin Psychiatry 58:278-282, 1997.

51. Fava M, Alpert JE, Nierenberg AA, Worthington JJ, Rosenbaum JF: Double-blind study of high dose fluoxetine versus lithium or desipramine augmentation of fluoxetine in partial and nonresponders to fluoxetine. Annual Meeting of the American Psychiatric Association. Chicago, Illinois, 2000

52. Fava M, Davidson KG. Definition and epidemiology of treatment-resistant depression. Psychiatr Clin North Am 19:179-200, 1996.

53. Fava M, Kaji J, Davidson K. Pharmacological strategies for treatment-resistant major depression. In Pollack MH, Otto MW, Rosenbaum JF, eds., Challenges in Clinical Practice: Pharmacological and Psychosocial Strategies. New York: Guilford Press, pp. 3-30, 1996.

54. Fava M, Kaji J. Continuation and maintenance treatments of major depressive disorder. Psychiatr Ann 24:281-290, 1994.

55. Fava M, Rosenbaum JF, Cohen L, Reiter S, McCarthy M, Steingard R, Clancy K. High-dose fluoxetine in the treatment of depressed patients not responsive to a standard dose of fluoxetine. J Affect Disord 25:229-234, 1992.

56. Pending publication
.submitted for publication.

57. Fava M, Rosenbaum JF, McGrath PJ, Stewart JW, Amsterdam JD, Quitkin FM. Lithium and tricyclic augmentation of fluoxetine treatment for resistant major depression: A double-blind, controlled study. Am J Psychiatry 151:1372-1374, 1994.

58. Fava M, Uebelacker LA, Alpert JE, Nierenberg AA, Pava J, Rosenbaum JF Major depressive subtypes and treatment response. Biol Psychiatry 42:568-576, 1997.

59. Francis A, Docherty JR, Kahn DA. The Expert Consensus Guideline Series. Treatment of Schizophrenia. J Clin Psychiatry 57(Suppl. 12a):1-58, 1996.

60. Francis A, Docherty JR, Kahn DA. The Expert Consensus Guideline Series: Treatment of Bipolar Disorder. J Clin Psychiatry 57(Suppl. 12b):1-88, 1996.

61. Frank E, Karp JF, Rush AJ. Efficacy of treatments for major depression. Psychopharmacol Bull 29:457-475, 1993.

62. Frank E, Kupfer DJ, Perel JM, Cornes C, Jarrett DB, Mallinger AG, Thase ME, McEachran AB, Grochocinski VJ. Three-year outcomes for maintenance therapies in recurrent depression. Arch Gen Psychiatry 47:1093-1099, 1990.

63. Frank E, Kupfer DJ, Siegel LR. Alliance not compliance: A philosophy of outpatient care. J Clin Psychiatry 56(Suppl 1):11-17, 1995.

64. Frank E, Perel JM, Mallinger AG, Thase ME, Kupfer DJ. Relationship of pharmacologic compliance to long-term prophylaxis in recurrent depression. Psychopharmacol Bull 28:231-235, 1992.

65. Frank E, Prien R, Jarrett RB, Keller MB, Kupfer DJ, Lavori P, Rush AJ, Weissman MM. Conceptualization and rationale for consensus definitions of response, remission, recovery, relapse and recurrence in major depressive disorder. Arch Gen Psychiatry 48:851-855, 1991.

66. Frazier-Smith N, Lesperance' F, Talajic M. Depression following myocardial infarction. Impact on six-months survival. JAMA 270:1819-1825, 1993.

67. Pending publication
(in press).

68. Friedman R, Myers P, Sobel D, Caudill M, Benson H. Behavioral medicine, clinical health psychology, and cost offset. Health Psychol 14:509-518, 1995.

69. Fulgum E, Rosenberg C, Damsbo N, Stage K, Lauritzen L, Bech P: Screening and treating depressed patients. A comparison of two controlled citalopram trials across treatment settings: Hospitalized patients versus patients treated by their family doctors. Danish University Antidepressant Group. Acta Psychiatr Scand 94:18-25, 1996.

70. Gemson DH, Ashford AR, Dickey LL, Raymore SH, Roberts JW, Ehrlich MH, Foster BG, Ganz ML, Moon-Howard J, Field LS, Bennett BA, Elinson J, Francis CK. Putting prevention into practice. Impact of a multifaceted physician education program on preventive services in the inner city. Arch Fam Med 155:2210-2216. 1995.

71. Gibbons RD, Hedeker D, Elkin I, Waternaux C, Kraemer HC, Greenhouse JB, Shea T, Imber SD, Sotsky SM, Watkins JT. Some conceptual and statistical issues in analysis of longitudinal psychiatric data.: Application to the NIMH Treatment of Depression Collaborative Research Program Dataset. Arch Gen Psychiatry 50:739-750, 1993.

72. Gilbert DA, Altshuler KZ, Rago WV, Shon SP, Crismon ML, Toprac MG, Rush AJ. Texas Medication Algorithm Project: Definitions, rationale and methods to develop medication algorithms. J Clin Psychiatry 59:345-351, 1998.

73. Gold MR, Siegel JE, Russell LB, Weinstein MC, eds. Cost-Effectiveness in Health and Medicine. New York: Oxford University Press, 1996.

74. Goldstein HI. Multilevel mixed linear model analysis using iterative generalized least squares. Biometrika 73:43-56, 1986.

75. Greenberg PE, Stiglin LE, Finkelstein SN, Berndt ER. The economic burden of depression in 1990. J Clin Psychiatry 54:405-418, 1993.

76. Grossman M. On the concept of health capital and the demand for health. J Polit Econ 80:223-255, 1972.

77. Guthrie E, Moorey J, Margison F, Barber H, Palmer S, McGrath G, Tomenson B, Creed F. Cost-effectiveness of brief psychodynamic-interpersonal therapy in high utilizers of psychiatric services. Arch Gen Psychiatry 56:519-526, 1999.

78. Gutierrez G, Guiscarfe H, Bronfman M, Walsh J, Martinez H, Munoz O. Changing physician prescribing patterns: Evaluation of an educational strategy for acute diarrhea in Mexico City. Med Care 32:436-446, 1994.

79. Guy W. ECDEU Assessment Manual for Psychopharmacology. Washington, DC: U.S. Government Printing Office, U.S. Dept. of Health, Education and Welfare Publication No. 76-338, 1976.

80. Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry 23:56-62, 1960.

81. Hamilton M. Development of a rating scale for primary depressive illness. Br J Soc Clin Psychol 6:278-296, 1967.

82. Hankin JR, Kessler LG, Goldberg ID, Steinwachs DM, Starfield BH. A longitudinal study of offset in the use of nonpsychiatric services following specialized mental health care. Med Care 21:1099-1110, 1983.

83. Hargreaves WA, Shumway M, Hu TW, Cuffel B. Cost-Outcome Methods for Mental Health. San Diego: Academic Press, 1998.

84. Headrick LA, Speroff T, Pelecanos HI, Cebul RD. Efforts to improve compliance with the national Cholesterol Education Program guidelines. Arch Intern Med 152:2490-2496, 1992.

85. Heckman J, Hotz VJ. Choosing among alternative non-experimental methods for estimating the impact of social programs: the case of manpower training. J Am Stat Assoc 84:862-874, 1989.

86. Heckman J. Instrumental variables: a study of implicit behavioral assumptions used in making program evaluations. J Human Resources 32:441-462, 1997.

87. Heckman J. Varieties of selection bias. Am Econ Rev 80:313-318, 1990.

88. Hedeker D. Random Regression Models with Autocorrelated Errors. Chicago, Ill. University of Chicago; 1989 Doctoral dissertation.

89. Hirschfeld RM, Keller MB, Panico S, Arons BS, Barlow D, Davidoff F, Endicott J, Froom J, Goldstein M, Gorman JM, Marek RG, Maurer TA, Meyer R, Phillips K, Ross J, Schwenk TL, Sharfstein SS, Thase ME, Wyatt RJ. The National Depressive and Manic-Depressive Association consensus statement on the undertreatment of depression. JAMA 277:333-340, 1997.

90. Holder HD, Blose JO. Changes in health care costs and utilization associated with mental health treatment. Hosp Comm Psychiatry 38:1070-1075, 1987.

91. Howland RH. General health, health care utilization, and medical comorbidity in dysthymia. Int J Psychiatr Med 23:211-238, 1993.

92. Hristov DH, Fallone BG. An active set algorithm for treatment planning optimization. Med Physics 24:1455-64, 1997.

93. Hunt DL, Haynes RB, Hanna SE, Smith K. Effects of computer-based clinical decision support systems on physician performance and patient outcomes. JAMA 280:1339-1346, 1998.

94. Institute of Medicine. Effectiveness and Outcomes in Health Care. Heithoff KA, Lohr KN (eds). Washington, D.C.: National Academy Press, 1990.

95. Jameson J, Shuman LJ, Young WW. The effects of outpatient psychiatric utilization on the costs of providing third party coverage. Med Care 16:383-399, 1978.

96. Jarrett RB, Rush AJ. Short-term psychotherapy of depressive disorders: Current status and future directions. Psychiatry: Interpers Biol Process 57:115-132, 1994.

97. Jennrich RI, Schluchter MD. Unbalanced repeated measures models with structured covariance matrices. Biometrics 42:805-820, 1986.

98. Jerrell JM, Hu T-W. Cost-effectiveness of intensive clinical and case management compared with an existing system of care. Inquiry 26:224-234, 1989.

99. Joffe RT, Levitt AJ, Sokolov ST, Young LT. Response to an open trial of a second SSRI in major depression. J Clin Psychiatry 57:114-115, 1996.

100. Joffe RT, Schuller DR. An open study of buspirone augmentation of serotonin reuptake inhibitors in refractory depression. J Clin Psychiatry 54:269-271, 1993.

101. Johnson DAW. A study of the use of antidepressant medication in general practice. Br J Psychiatry 125:186-192, 1974.

102. Johnson GF. Lithium in depression: A review of the antidepressant and prophylactic effects of lithium. Aust NZ J Psychiatry 21:356-365, 1987.

103. Johnson J, Weissman M, Klerman G. Service utilization and social morbidity associated with depressive symptoms in the community. JAMA 267:1478-1483, 1992.

104. Jones KR, Vischi TR. Impact of alcohol, drug abuse and mental health treatment on medical care utilization. M Care 12(Suppl), 1979.

105. Judd LL, Akiskal HS, Maser JD, Zeller PJ, Endicott J, Coryell W, Paulus MP, Kunovac JL, Leon AC, Mueller TI, Rice JA, Keller MB. A prospective 12-year study of subsyndromal and syndromal depressive symptoms in unipolar major depressive disorders. Arch Gen Psychiatry 55:694-700, 1998.

106. Judd LL, Paulus MP, Wells KB. Socioeconomic burden of subsyndromal depressive symptoms and major depression in a sample of the general population. Am J Psychiatry 153:1411-1416, 1996.

107. Judd LL. The clinical course of unipolar major depressive disorders. Arch Gen Psychiatry 54:989-991, 1997.

108. Kashner TM, Rost K, Smith GR, Lewis S. An analysis of panel data: The impact of a psychiatric consultation letter on the expenditures and outcomes of care for patients with somatization disorder. Med Care 30:811-821, 1992.

109. Kashner TM, Rush AJ, Altshuler KZ. Managed care and the focus on outcomes research. J Practical Psychiatry Behav Health 3:135-145, 1997.

110. Kashner TM, Rush AJ, Altshuler KZ. Measuring costs of guideline-driven mental health care: The Texas Medication Algorithm Project. J Ment Health Policy Econ 2:111-121, 1999.

111. Kashner TM, Rush AJ. Measuring cost offsets of psychotherapy. In Miller NE, Magruder KM, eds., The Cost Effectiveness of Psychiatry: A Guide for Practitioners, Researchers, and Policymakers. New York: Oxford University Press, 1999.

112. Kashner TM, Shoaf T, Rush AJ. The economic burden of suicide in the United States in the Year 2000. Econ Neurosci 2: 44-48, 2000.

113. Kashner TM, Suppes T, Rush AJ, Altshuler KZ. Measuring use of outpatient care among mentally ill individuals: A comparison of self reports and provider records. Eval Prog Plann 22:31-39, 1999.

114. Kashner TM. Agreement between administrative files and written medical records: A case of the Department of Veterans Affairs. Med Care 36:1324-1336, 1998.

115. Kashner TM. Cost effectiveness of lead screening, comment. N Eng J Med 307:1260, 1982.

116. Keitner GI, Ryan CE, Miller IW, Norman WH. Recovery and major depression: Factors associated with twelve-month outcome. Am J Psychiatry 149:93-99, 1992.

117. Keller MB, Gelenberg AJ, Hirschfeld RMA, Rush AJ, Thase ME, Kocsis JH, Markowitz JC, Fawcett JA, Koran LM, Klein DN, Russell JM, Kornstein SG, McCullough JP, Davis SM, Harrison WM. The treatment of chronic depression, Part 2: A double-blind, randomized trial of sertraline and imipramine. J Clin Psychiatry 59:598-607, 1998.

118. Keller MB, Harrison W, Fawcett JA, Gelenberg A, Hirschfeld RMA, Klein D, Kocsis J, McCullough JP, Rush AJ, Schatzberg A, Thase ME. Treatment of chronic depression with sertraline or imipramine: Preliminary blinded response rates and high rates of undertreatment in the community. Psychopharmacol Bull 31:205-212, 1995.

119. Keller MB, Klerman GL, Lavori PW, Fawcett JA, Coryell W, Endicott J. Treatment received by depressed patients. JAMA 1848-1855, 1982.

120. Kelsey JE. Dose-response relationship with venlafaxine. J Clin Psychopharmacol 16(Suppl. 2):21S-28S, 1996.

121. Kessler LG, Burns BJ, Shapiro S, Tischler GI, George LK, Hough RL, Bodison D, Miller RH. Psychiatric diagnoses of medical service users: evidence from the Epidemiologic Catchment Area program. Am J Public Health 77:18-24, 1987.

122. Kessler RC, McGonagle KA, Zhao S, Nelson CB, Hughes M, Eshleman S, Wittchen H-U, Kendler KS. Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Results from the National Comorbidity Survey. Arch Gen Psychiatry 51:8-19, 1994.

123. Ketai R. Family practitioners knowledge about treatment of depressive illness. JAMA 235:2600-2603, 1976.
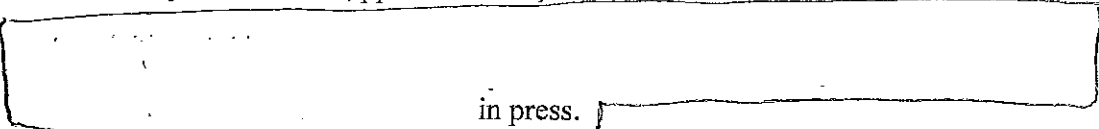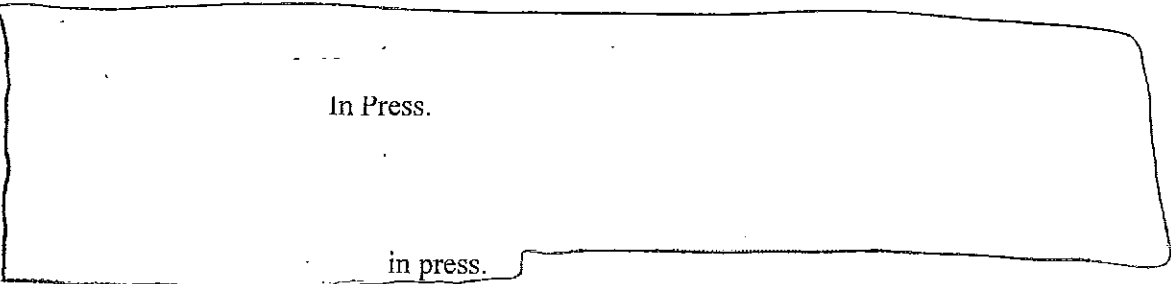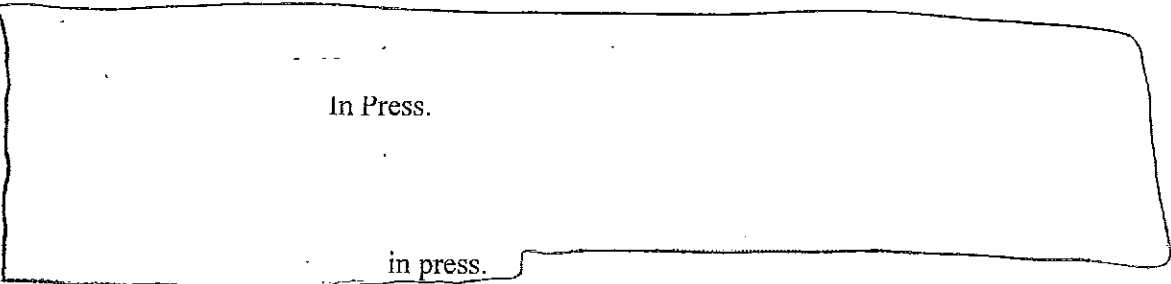
124. Kmenta J. Elements of Econometrics. New York: MacMillan, 1973.

125. Kobak KA, Greist JH, Jefferson JW, Mundt JC, Katzelnick DJ. Computerized assessment of depression and anxiety over the telephone using Interactive Voice Response. MD Computing May/June:61-68, 1999.

126. Kobak KA, Greist JH, Jefferson JW, Mundt JC, Katzelnick DJ. Computer-administered clinical rating scales: A review. Psychopharmacology 127:291-300, 1996.

127. Kobak KA, Reynolds WM, Rosenfeld R, Greist JH. Development and validation of a computer-administered version of the Hamilton Depression Rating Scale. Psychol Assess. J Consult Clin Psychol 2:56-63, 1990.

128. Kobak KA, Reynolds WM, Greist JH. Development and validation of a computer-administered version of the Hamilton Anxiety Scale. Psychol Assess 5:487-492, 1993.

129. Kotin J, Post RM, Goodwin FK. Drug treatment of depressed patients referred for hospitalization. Am J Psychiatry 130:1139-1141, 1973.

130. Krumholz HM, Pasternak RC, Weinstein MC, et al. Cost effectiveness of thrombolytic therapy with streptokinase in elderly patients with suspected acute myocardial infarction. N Eng J Med 327:7-13, 1992.

131. La Ruche G, Lorougnon F, Digbeu N. Therapeutic algorithms for the management of sexually transmitted diseases at the peripheral level in Côte d'Ivoire: assessment of efficacy and cost. Bulletin of the World Health Organization 73:305-313, 1995.

132. Laird NM, Ware H. Random-effects models for longitudinal data. Biometrics 38:963-974, 1982.

133. Lave JR, Frank RG, Schulberg HC, Kamlet MS. Cost-effectiveness of treatments for major depression in primary care practice. Arch Gen Psychiatry 65:645-651, 1998.

134. Levenson J, Hamer RM, Rossiter LF. Relation of psychopathology in general medical inpatients to use and cost of services. Am J Psychiatry 147:1498-1503, 1990.

135. Linn BS, Linn MW, Gurel L. Cumulative Illness Rating Scale. J Am Geriatr Soc 16:622-626, 1968.

136. Louis TA, Spiro A, III. Fitting first order auto-regressive models with covariates (Technical Report). Cambridge, MA: Harvard University School of Public Health, Department of Biostatistics, 1984.

137. Lubarsky DA, Glass PSA, Ginsberg B, Dear G de L, Dentz ME, Gan TJ, Sanderson IC, Mythen MG, Dufore S, Pressley C, Gilbert WC, White WD, Alexander ML, Coleman RL, Rogers M, Reves JG. The successful implementation of pharmaceutical practice guidelines: analysis of associated outcomes and cost savings. Anesthesiology 86:1145-1160, 1997.

138. Maddala GS. Limited-Dependent and Qualitative Variables in econometrics. New York: Cambridge University Press, 1988.

139. Magaziner J, Simonsick EM, Kashner TM, Hebel JR. Patient-Proxy response comparability on measures of patient health and functional status. J Clin Epidemiol 41:1065-1074, 1988.

140. Magaziner J, Simonsick EM, Kashner TM, Hebel JR. Predictors of functional recovery one year following hospital discharge for hip fracture: a prospective study. J Gerontol: Med Sciences 45:M101-107, 1990.

141. Manning W, Wells KB. The effects of psychological distress and psychological well-being on use of medical services. Med Care 30:541-553, 1992.

142. Manski CF. Learning about treatment effects from experiment with random assignment of treatments. J Human Resources 31:707-733, 1996.

143. Margolis CZ, Warshawsky SS, Goldman L, Dagan O, Wirtschafter D, Pliskin JS. Computerized algorithms and pediatricians' management of common problems in a community clinic. Academic Med 67:282-284, 1992.

144. Maritz TS. Empirical Bayes Methods. London (England): Methuen, 1970.

145. Marshall RD, Liebowitz MR. Paroxetine/bupropion combination treatment for refractory depression (letter). J Clin Psychopharmacol 16:80-81, 1996.

146. Mason BJ, Markowitz JC, Klerman GL. Interpersonal psychotherapy for dysthymic disorders. In Klerman GL, Weissman MM, eds., New Applications of Interpersonal Psychotherapy. Washington, DC: American Psychiatric Press, Inc., pp. 225-264, 1993.

147. McCombs JS, Nichol MB, Stimmel GL, Sclar DA, Beasley CM Jr, Gross LS. The cost of antidepressant drug therapy failure: A study of antidepressant use patterns in a Medicaid population. J Clin Psychiatry 51(Suppl.):60-69, 1990.

148. McCombs JS, Nichol MB. Pharmacy-enforced outpatient drug treatment protocols: A case study of Medi-Cal restrictions for cefaclor. Ann Pharmacother 27:155-161, 1993.

149. McFadden Jr ER, Elsanadi N, Dixon L, Takacs M, Deal EC, Boyd KK, Idemoto BK, Broseman LA, Panuska J, Hammons T, Smith B, Caruso F, McFadden CB, Shoemaker L, Warren EL, Hejal R, Strauss L, Gilbert IA. Protocol therapy for acute asthma: therapeutic benefits and cost savings. Am J Med 99:651-661, 1995.

150. McGrath PJ, Stewart JW, Nunes EN, Quitkin FM. Treatment response of depressed outpatients unresponsive to both a tricyclic and a monoamine oxidase inhibitor antidepressant. J Clin Psychiatry 55:336-339, 1994.

151. McGrath PJ, Stewart JW, Nunes EV, Ocepek-Welikson K, Rabkin JG, Quitkin FM, Klein DF. A double-blind crossover trial of imipramine and phenelzine for outpatients with treatment-refractory depression. Am J Psychiatry 150:118-123, 1993.

152. McHugh JP, Kahn MW, Heiman E. Relationships between mental health treatment and medical care utilization among low-income Mexican-American patients: some preliminary findings. Med Care 15:439-444, 1977.

153. Mechanic D, Cleary PD, Greenley JR. Distress syndromes, illness behavior, access to care and medical utilization in a defined population. Med Care 20:361-372, 1982.

154. Menke TJ, Homan RH, Kashner TM. Issues in determining costs in the Department of Veterans Affairs. Med Care 4:AS18-AS26, 1999.

155. Meterissian GB, Bradwejn J. Comparative studies on the efficacy of psychotherapy, pharmacotherapy, and their combination in depression: Was adequate pharmacotherapy provided? J Clin Psychopharmacol 9:334-339, 1989.

156. Miller IW, Keitner GI, Schatzberg AJ, Klein DN, Thase ME, Rush AJ, Markowitz JC, Schlager DS, Kornstein SG, Davis SM, Harrison WM, Keller MB The treatment of chronic depression, Part 3: Psychosocial functioning before and after treatment with sertraline or imipramine. J Clin Psychiatry 59:608-619, 1998.

157. 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 41:237-248, 1992.

158. Miller MD, Towers A. A manual of guidelines for scoring the cumulative rating scale for geriatrics (CIRS-G). University of Pittsburgh, May 1991.

159. Miller NE, Magruder KM (eds). The Cost Effectiveness of Psychotherapy: A Guide for Practitioners Researchers and Policymakers. New York: Oxford University Press, 1999.

160. Mishan EJ. Cost-Benefit Analysis. New York: Praeger Publishers, 1976.

161. Morris CB, Rittenhouse BE (eds.) Standards for Economic Evaluation of Drugs: Issues and Answers. Med Care 34:DS1-DS239, 1996.

162. Mossey JM, Mutran E, Knott K, Craik R. Determinants of recovery 12 months after hip fracture: the importance of psychosocial factors. Am J Public Health 79:279-286, 1989.

163. Mueller TI, Leon AC. Recovery, chronicity, and levels of psychopathology in major depression. Psychiatr Clin North Am 19:85-102, 1996.

164. Mumford E, Schlesinger HJ, Glass GV, Patrick C, Cuerdon T. A new look at evidence about reduced cost of medical utilization following mental health treatment. Am J Psychiatry 141:1145-1158, 1984.

165. Mundt JC, Kobak KA, Greist JH, Marks IM. Work and Social Adjustment Scale: An Alternative Quality-of-Life Instrument.

166. Mundt J. Interactive voice response systems in clinical research and treatment. Psychiatric Serv 48:611-623, 1997.

167. Murray CJ, Lopez, AD. Evidence-based health policy — Lessons from the Global Burden of Disease Study. Science 274:740-743, 1996.

168. NIMH. The Future of Mental Health Services Research. Taube CA, Mechanic D, Hohmann A (eds). DHHS Pub No (ADM)89-1600. Washington, D.C.: US Govt Printing Office, 1989.

169. Nierenberg AA, Feighner JP, Rudolph R, Cole JO, Sullivan J. Venlafaxine for treatment-resistant unipolar depression. J Clin Psychopharmacol 14:419-423, 1994.

170. Nierenberg AA, McLean NE, Alpert JE, Worthington JJ, Rosenbaum JF, Fava M. Early nonresponse to fluoxetine as a predictor of poor 8-week outcome. Am J Psychiatry 152:1500-1503, 1995.

171. Nierenberg AA, Mulroy R. Declaration of treatment failures. In Rush AJ, ed., Modern Problems of Pharmacopsychiatry, Vol. 25. Mood Disorders. Systematic Medication Management, Basel: S Karger, pp. 17-33, 1997.

172. Nierenberg AA, Price LH, Charney DS, Heninger GR. After lithium augmentation: A retrospective follow-up of patients with antidepressant-refractory depression. J Affect Disord 18:167-175, 1990.

173. Ontivarios A, Fontaine R, Elie R. Refractory depression: The addition of lithium to fluoxetine or desipramine. Acta Psychiatr Scand 83:188-192, 1991.

174. Oquendo MA, Malone KM, Mann JJ. Suicide: Risk factors and prevention in refractory major depression. Depression and Anxiety 5:202-211, 1997.

175. Oquendo MA, Malone KM, Ellis SP, Sackeim HA, Mann JJ. Antidepressant treatment is inadequate in patients with major depression who are at risk for suicidal behavior. Am J Psychiatry 156:190-194, 1999.

176. Patrick DL, Erickson P. Health Status and Health Policy: Allocating Resources to Health Care. New York: Oxford University Press, 1993.

177. Patterson D, Bise B. Report to NIMH, Contract 282-77-0219 MS. Rockville, MD, 1978.

178. Paykel ES, Hollyman JA, Freeling P, Sedgwick P. Predictors of therapeutic benefit from amitriptyline in mild depression: A general practice placebo-controlled trial. J Affect Disord 14:83-95, 1988.

179. Paykel ES, Klerman GL, Prusoff BA. Treatment setting and clinical depression. Arch Gen Psychiatry 22:11-21, 1970.

180. Paykel ES, Scott J, Teasdale JD, Johnson AL, Garland A, Moore R, Jenaway A, Cornwall PL, Hayhurst H, Abbott R, Pope M. Prevention of relapse in residual depression by cognitive therapy. A controlled trial. Arch Gen Psychiatry 1999;56:829-835.

181. Pohlmeier W, Ulrich V. An econometric model of the two-part process in the demand for health. J Human Resources 30:339-361, 1995.

182. Preskorn SH, Lane RM. Sertraline 50 mg daily; the optimal dose in the treatment of depression. Int Clin Psychopharmacol 10:129-141, 1995.

183. Quam L, Ellis LBM, Venus P, Clouse J, Taylor CG, Leatherman S. Using claims data for epidemiologic research: The concordance of claim-based criteria with the medical record and patient survey for identifying a hypertensive population. Med Care 31:498-507, 1993.

184. Quitkin FM, Rabkin JG, Ross D, McGrath PJ. Duration of antidepressant drug treatment. What is an adequate trial? Arch Gen Psychiatry 41:238-245, 1984.

185. Ravaris CL, Nies A, Robinson D et al. A multiple dose controlled study of phenelzine in depression-anxiety states. Arch Gen Psychiatry 33:347-350, 1976.

186. Reilly MC, Tanner A, Meltzer EO. Work, classroom and activity impairment instruments validation studies in allergic rhinitis. Clin Pharmacoeconomics. 4 (5) 1993.

187. Remick RA, Barton JS, Patterson B. On so-called treatment-resistant depression. Presented at 51st Annual Meeting of the Royal Collage of Physicians and Surgeons of Canada, Québec City, Québec, Canada, September 1982.

188. Rice DP, Hodgson TA, Kapstein AN. The economic costs of illness: A replication and update. Health Care Fin Rev 7:61-80, 1985.

189. Rice DP, Miller LS. Economic burden of affective disorders. In Hu T-W, Rupp A eds. Advances in Health Economics and Health Services Research. Vol. 14. Greenwich, Connecticut: JAI Press, 1993.

190. Rice DP. Economic burden of mental disorders in the United States. Econ Neurosci 1:40-44, 1999.

191. Riso LP, Thase ME, Howland RH, Friedman ES, Simons AD, Tu XM. A prospective test of criteria for response, remission, relapse, recovery, and recurrence in patients treated with cognitive behavior therapy. J Affect Disord 43:131-142, 1997.

192. Roland PY, Naumann RW, Alvarez RD, Kilgore LC, Partridge EE. A decision analysis of practice patterns used in evaluating and treating abnormal pap smears. Gynecol Oncol 59:75-80, 1995.

193. Rosenbaum JF, Fava M, Nierenberg AA, Sachs G. Treatment-resistant mood disorders. In Gabbard GO, ed., Treatments of Psychiatric Disorders: Second Edition. Washington, DC: American Psychiatric Press, pp. 1275-1328, 1995.

194. in press.

195. Rupp A. The economic consequences of not treating depression. Br J Psychiatry 166 (Suppl27):29-33, 1995.

196. In Press.

197. in press.

198. Rush AJ, Giles DE, Schlesser MA, Fulton CL, Weissenburger JE, Burns CT. The Inventory for Depressive Symptomatology (IDS): Preliminary findings. Psychiatry Res 18:65-87, 1986.

199. Rush AJ, Gullion CM, Basco MR, Jarrett RB, Trivedi MH. The Inventory of Depressive Symptomatology (IDS): Psychometric properties. Psychol Med 26:477-486, 1996.

200. Rush AJ, Kupfer DJ. Strategies and tactics in the treatment of depression. In Gabbard GO, ed., Treatments of Psychiatric Disorders, Second Edition, Volume 1. Washington, DC: American Psychiatric Press, Inc., pp. 1349-1368, 1995.

pending publications

201. Rush AJ, Prien RF. From scientific knowledge to the clinical practice of psychopharmacology: Can the gap be bridged? Psychopharmacol Bull 31:7-20, 1995.

202. Rush AJ, Rago WV, Crismon ML, Toprac MG, Shon SP, Suppes T, Miller AL, Trivedi MH, Swann AC, Biggs MM, Shores-Wilson K, Kashner TM, Pigott T, Chiles JA, Gilbert DA, Altshuler KZ. Medication treatment of the severely and persistently mentally ill: The Texas Medication Algorithm Project. J Clin Psychiatry 60:142-156, 1999.

203. Rush AJ, Thase ME. Indications and planning of psychotherapies. In Maj M, Sartorius N, eds., World Psychiatric Association Series in Evidence and Practice in Psychiatry. Volume 1: Depressive Disorders. Chichester, UK: John Wiley & Sons, Ltd., 1999.

204. Rush AJ, Trivedi MH. Treating depression to remission. Psychiatr Ann 25:704-705, 709, 1995.

205. Russell LB, Gold MR, Siegel JE, Daniels, N, Weinstein MC. The role of cost-effectiveness analysis in health and medicine. JAMA 276:1172-1177, 1996.

206. Salvador-Carulla L, Segui J, Fernandez-Cano P, Canet J. Costs and offset effect in panic disorder. Br J Psychiatry 166 (Supplement 27):23-28, 1995.

207. Schatzberg AF, Cole JO, Cohen BM, et al Survey of depressed patients who failed to respond to treatment. In Davis JM, Maas JW, eds., The Affective Disorders. Washington, DC: American Psychiatric Press, pp. 73-85, 1983.

208. Schoenbaum SC, Gottlieb LK. Algorithm based improvement of clinical quality. BMJ 301:1374-1376, 1990.

209. Schulman KA, Yabroff KR. Measuring the cost-effectiveness of cancer care. Oncology 9:523-538, 1995.

210. Schweizer E, Rickels K, Amsterdam JD, Fox I, Puzzuoli G, Weise C. What constitutes an adequate antidepressant trial for fluoxetine? J Clin Psychiatry 51:8-11, 1990.

211. Shtasel DL, Gur RE, Mozley PD, Richards J, Taleff MM, Heimberg C, Gallacher F, Gur RC. Volunteers for biomedical research. Recruitment and screening of normal controls. Arch Gen Psychiatry 48:1022-1025, 1991.

212. Siegel JE, Laska E, Meisner M. Statistical methods for cost-effectiveness analyses. Controlled Clin Trials 17:387-404, 1996.

213. Siegel JE, Weinstein MC, Russell LB, Gold MR. Recommendations for reporting cost-effectiveness analyses. JAMA 276: 1339-1341, 1996.

214. Simon GE, Katzelnick DJ. Depression, use of medical services and cost-offset effects. J Psychosom Res 42:333-344, 1997.

215. Simon GE, Von Korff M, Barlow W. Health care costs of primary care patients with recognized depression. Arch Gen Psychiatry 52:850-856, 1998.

216. Simpson GM, Lee JH, Cuculica A, Kellner R. Tow dosages of imipramine in hospitalized endogenous and neurotic depressives. Arch Gen Psychiatry 33:1093-1102, 1976.

217. Stanek III EJ, Diehl SR. Growth curve models of repeated binary response. Biometrics 44:973-983, 1988.

218. Stassen HH, Delini-Stula A, Angst J. Time course of improvement under antidepressant treatment: A survival-analytic approach. Eur Neuropsychopharmacol 3:127-135, 1993.

219. Stoudemire A, Frank R, Hedemark N, Kamlet M, Blazer D. The economic burden of depression. Gen Hosp Psychiatry 8:387-394, 1986.

220. Taube CA, Kessler LG, Burns BJ. Estimating the probability and level of ambulatory mental health services use. Health Serv Res 21: 321-340, 1986.

221. Thase ME, Blomgren SL, Birkett MA, Apter JT, Tepner RG. Fluoxetine treatment of patients with major depressive disorder who failed initial treatment with sertraline. J Clin Psychiatry 58:16-21, 1997.

222. Thase ME, Howland RH. Refractory depression: Relevance of psychosocial factors and therapies. Psychiatr Ann 24:232-240, 1994.

223. Thase ME, Lydiard RB, Feighner JP. Citalopram treatment of fluoxetine nonresponders. APA Annual Meeting, Washington, DC, 1999.

224. Thase ME, Mallinger AG, McKnight D, Himmelhock HM. Treatment of imipramine-resistant recurrent depression, IV: A double-blind crossover study of tranylcypromine for anergic bipolar depression. Am J Psychiatry 149:195-198, 1992.

225. Thase ME, Rush AJ. When at first you don't succeed: Sequential strategies for antidepressant nonresponders. J Clin Psychiatry 58(Suppl. 13):23-29, 1997.

226. Thase ME, Rush AJ: Treatment-resistant depression. In: Bloom FE, Kupfer DJ, eds. Psychopharmacology: Fourth Generation of Progress. New York: Raven Press, pp. 1081-1097, 1995.

227. Theil H. Principles of Econometrics. NY: Wiley, 1971.

228. Tierney WM, Hui SL, McDonald CJ. Delayed feedback of physician performance versus immediate reminders to perform preventive care. Effects on physician compliance. Med Care 24:659-666, 1986.

229. Torrance GW, Feeny D. Utilities and Quality-Adjusted Life Years. Intl J Tech Assess Health Care 5:559-575, 1989.

230. Tyrer P, Gardner M, Lambourn J, Whitford M. Clinical and pharmacokinetic factors affecting response to phenelzine. Br J Psychiatry 136:359-365, 1980.

231. van Praag, H., Asnis GM, Kahn RS, Brown, SL, Korn M, Friedman JM, Wetzler S. Monoamines and abnormal behaviour. A multi-aminergic perspective. Br J Psychiatry 157:723-734, 1990.

232. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 30:473-483, 1992.

233. Ware JH. Linear models for the analysis of longitudinal studies. Am Statistician 39:95-101, 1985.

234. Warner KE, Luce BR. Cost-Benefit and Cost Effectiveness in Health Care. Ann Arbor: Health Administration Press, 1982.

235. Watt DC, Crammer JL, Elkes A. Metabolism, anticholinergic effects and therapeutic effects on outcome of desmethylimipramine in depressive illness. Psychol Med 2:397-405, 1972.

236. Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russell LB. Recommendations of the panel on cost-effectiveness in health and medicine. JAMA 276:1253-1258, 1996.

237. Weinstein MC, Stason WB. Foundations of cost-effectiveness analysis for health and medical practices. N Eng J Med 296:716-721, 1977.

238. Weissman MM, Bruce ML, Leaf PJ, Florio LP, Holzer C III. Affective disorders. In Robins LN, Regier DA, eds., Psychiatric Disorders in America: The Epidemiologic Catchment Area Study. New York: The Free Press, pp. 53-80, 1991.

239. Wells KB, Stewart A, Hayes RD, Burnam MA, Rogers W, Daniels M, Berry S, Greenfield S, Ware J. The functioning and well-being of depressed patients: Results from the Medical Outcomes Study. JAMA 262:914-919, 1989.

240. Wells KB, Sturm R, Sherbourne CD, Meredith LS. Caring for Depression. Cambridge (MA): Harvard University Press, 1996.

241. Williams JB. A structured interview guide for the Hamilton Depression Rating Scale. Arch Gen Psychiatry 45:742-747, 1988.

242. Willett JB, Ayoub CC, Robinson D. Using growth modeling to examine systematic differences in growth: an example of change in the functioning of families at risk of maladaptive parenting, child abuse, or neglect. J Consult Clin Psychol 59:38-47, 1991.

243. Yates B. Cost-effectiveness analysis, cost-benefit analysis, and beyond. Evolving models for the scientist-manager-practitioner. Clin Psychol: Sci Pract 2:385-398, 1995.

244. Zhang M, Rost KM, Fortney JC. Earnings changes for depressed individuals treated by mental health specialists. Am J Psychiatry 156:108-114, 1999.

245. Zimmerman M, Mattia JI. The reliability and validity of a screening questionnaire for 13 DSM-IV Axis I disorders (the Psychiatric Diagnostic Screening Questionnaire) in psychiatric outpatients. J Clin Psychiatry 60:677-683, 1999.