TCU School of Medicine

Historical Data Retrieval in Relation to Screen Failure Rate at Central Rater Call: A Retrospective Comparative Analysis of Three Clinical Trials

Scholarly Pursuit & Thesis

Morgan Moses
2-28-2023

Abstract

Research Question

What is the effect of central rater interviews on screen failure rate in clinical trials for major depressive disorder compared to screen failure rate at on-site interview?

Background, Significance, and Rationale for the Question

MDD is one of the most common mental health conditions in the United States. Currently available treatments for MDD are often inadequate, although research into new treatments is ongoing. Clinical trials that investigate new medications to treat MDD often employ central raters to screen candidates in order to increase the internal validity of their study. Central raters remotely evaluate the candidate to determine whether or not they are eligible for the study. Central rater evaluation of MDD severity may differ from on-site evaluation leading to a higher rate of screen failures.

Materials and Methods

Retrospective analysis of the demographics of candidates that failed screening and the reason for their ineligibility in three clinical trials at one clinical site. Historical data retrieval conducted for each person screened.

Anticipated Results, Conclusions, and Impact

We anticipate that screen failures will be higher in clinical trials for major depressive disorder that employ central raters.

Research Question

What is the effect of central rater interviews on screen failure rate in clinical trials for major depressive disorder compared to on-site interviews?

Hypothesis

We hypothesize that screen failures will be higher in clinical trials for major depressive disorder that employ central raters.

Introduction

Major depressive disorder (MDD) is one of the most common mental health disorders in the United States with 20.6% of noninstitutionalized adults having experienced the condition as defined by

the DSM-5 at some point in their life, while 10.4% reported experiencing the condition within the previous 12 months.[1] Characterized by emotional, cognitive, and somatic disturbances, the condition is strongly associated with decreased satisfaction with life for those affected, significantly decreased daily functioning, and social withdrawal.[1,2] In fact, a study conducted found that depressive disorders was one of the leading global causes of "years lived with disability" and "disability adjusted life years" in 2010, even without adjusting for disease processes attributable to depressive disorders. This signifies the large role that depressive disorders play in disease burden, despite being considered non-fatal. Furthermore, MDD accounted for 85% of "years lived with disability" and "disability adjusted life years" due to depressive disorders, emphasizing the intense disease burden of MDD.[3]

The DSM-5 describes nine symptoms of major depressive disorder and asserts that five or more of the symptoms must have been present during the same 2-week period and demonstrate a change in previous functioning. The symptoms can be organized into three categories: emotional disturbances, disturbances of cognition, or somatic disturbances. At least one of the symptoms associated with emotional disturbances must be present during the same 2-week period in order for a diagnosis to be considered. The symptoms present must cause clinically significant distress or impairment in social, occupational, or other important areas of functioning, and often number of symptoms present are correlated to increased disease severity and decreased functioning. Moreover, none of the symptoms can be clearly attributed to another medical condition or the effects of a substance. Exclusion of other mental health conditions that may better explain the occurrence of a major depressive episode is necessary before diagnosing MDD. Other mental health conditions include schizoaffective disorder, schizophrenia, delusional disorder, other psychotic disorders, or bipolar disorder.[4,5]

The symptoms of depressed mood (dysphoria) or loss of interest or pleasure (anhedonia) represent emotional disturbances that may take place in someone who has MDD. Dysphoria may manifest in a variety of ways, including feeling discouraged, numb, or anxious. Irritability, anger, persistent annoyance, hostility, or frustration may occur in up to 50% of MDD patients.[6] Anhedonia can cause someone to lose interest or pleasure in activities that they previously enjoyed, as well as withdraw from others.[4,5]

Disturbances in cognition may manifest in symptoms such as feelings of worthlessness or excessive or inappropriate guilt, recurrent thoughts of death, suicidal ideation and suicidal behavior, and diminished ability to think or concentrate or indecisiveness. One meta-analysis revealed that people with depression demonstrated "significant moderate cognitive deficits in cognitive function, memory and attention" in comparison to controls.[7] Another meta-analysis verified such claims, adding that theory of mind, or the ability to think about the mental state of yourself or others, is significantly affected in those with MDD in proportion to the severity of the symptoms.[8] In fact, direct reports from patients with MDD

demonstrated that difficulty concentrating and indecisiveness were two of the most distressing symptoms while issues with executive function and working memory are the most frequently reported.[9] When dysphoria or anhedonia was at its worst in those with lifetime MDD, "34.8% thought about their own death, 46.7% wanted to die, and 39.3% contemplated suicide," while in those who had experienced MDD within the last 12 months, 28.8% thought about their own death, 32.1% wanted to die, and 22.8% contemplated suicide.[1]

Somatic disturbances account for the remaining four symptoms: changes in weight or appetite, sleep disturbance, fatigue or loss of energy (anergia), or psychomotor agitation or retardation. Sleep disturbances can include insomnia or hypersomnia, which may significantly affect daily functioning. Anergia can present as a feeling of tiredness, exhaustion, or heaviness. Patients may require rest throughout the day or may make it difficult to start or complete tasks. Psychomotor disturbances involving agitation manifest as excessive movement that is often nonproductive and repetitive, such as fidgeting, whereas disturbances involving retardation manifest as slowness of movements, speech, or thought.[4,5] Along with changes in cognition, a change in psychomotor speed is one of the most frequently reported symptoms of MDD.[9]

A number of factors make a person more at risk for MDD, including experiencing a chronical medical illness, chronic minor daily stress, chronic pain syndrome, family history of depression, female sex, low income or job loss, low self-esteem, low social support, prior depression, being single, divorced or widowed, traumatic brain injury, and younger age.[10] Inspecting some of these risk factors in depth, such as sex, age, race, and income demonstrates the extent of the discrepancies between groups. Among women, the lifetime prevalence of MDD is 26.1% and the 12-month prevalence is 13.4%. Among men, however, the lifetime prevalence of MDD is 14.7% and the 12-month prevalence is 7.2%. This data suggests that the prevalence of MDD among women is nearly twice as high as among men. The 12-month prevalence of MDD is greater in age groups younger than 65 years of age, while the mean age of onset is 29 years of age and the mean age at first treatment is 32 years of age. In terms of race, MDD has a greater prevalence in white and Native American adults than African American, Asian American, and Hispanic adults. Although there has been found to be little difference above a $70,000 income, the odds of MDD increase with decreased household income.[1]

When a disease has such a significant impact, both in terms of symptoms and prevalence, treatment is crucial to ease the burden of the disease. Yet, 30% of those with a lifetime prevalence of MDD are untreated. Out of those who seek treatment, only 47% patients respond to standard antidepressant treatment, 50% of patients will experience a relapse or chronic form of MDD, and some studies report that the remission rate after treatment standard antidepressant treatment is only 33%. Furthermore, the time antidepressant takes to reach a noticeable effect is usually at least two weeks,

which is a significant amount of time for a patient to wait for relief from a disease. [1,11,13,14] Such statistics demonstrate that the existing treatment for MDD is not adequate to meet the needs of those affected by MDD. In fact, pharmaceutical innovation in MDD treatment has nearly failed to extend past the monoamine hypothesis of depression.

In the 1950s, the formulation of the monoamine hypothesis of depression laid the foundation for the discovery the first pharmacological treatment for MDD. The monoamine hypothesis contended that the pathophysiological basis of depression was the decreased concentrations of monoamines such as serotonin, noradrenaline, and dopamine in the synaptic gaps of those with depression. Eight years later, the first successful pharmacological treatment for depression was released, although it was marketed as a tuberculosis treatment. Isoniazid, a monoamine oxidase inhibitor (MAOI), inhibited monoamine oxidase enzymes from breaking down monoamines in the presynaptic cleft, therefore increasing the monoamine concentration available. The next year, imipramine, a tricyclic antidepressant (TCA), was approved for the treatment of MDD. The mechanism of action of TCAs differ slightly from that of MAOIs but follows the same principle. TCAs act by inhibiting the reuptake transporter proteins that clear norepinephrine and serotonin in the synaptic cleft, therefore increasing the available concentrations of these monoamines. TCAs differ from MAOIs, however, because the medication also blocks postsynaptic adrenergic α1 and α2 receptors, postsynaptic muscarinic receptors, and postsynaptic histamine H1 receptors. While whispers of the role of serotonin in MDD began in the late 1960s, the first selective serotonin reuptake inhibitor (SSRI) was not approved for the treatment of MDD until the later part of 1987. Similar to the action of TCAs, SSRIs inhibit the transporter proteins responsible for the reuptake of serotonin the synaptic cleft. SSRIs are markedly different than TCAs beyond this action, however, as this class of medication demonstrates a significant selectivity for serotonin over norepinephrine without the same binding affinity for postsynaptic adrenergic α1 and α2 receptors, postsynaptic muscarinic receptors, or postsynaptic histamine H1 receptors. While other drugs have been developed since, such as bupropion a year later in 1989, venlafaxine in 1993, and most recently vortioxetine in 2013, they have similar mechanisms of actions to the antidepressant medications already on the market.[11,12]

In clinical practice, initial MDD treatment and management has been generally governed by a stepwise treatment algorithm proposed by the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study. The first step is monotherapy treatment with a SSRI class antidepressant medication, specifically citalopram. If the medication is not tolerated by the patient or does not cause remission of their symptoms, then the patient will move to the next step. The next stage directs the patient to switch to sustained release bupropion, cognitive therapy, a different SSRI, or extended-release venlafaxine. Another option would be to augment the initial treatment with sustained-release bupropion, buspirone, or cognitive therapy. The third stage suggests the patient switch to mirtazapine or nortriptyline, or augment

treatment with lithium or nortriptyline (a tricyclic antidepressant; if the treatment is sustained-release bupropion or extended-release venlafaxine). The last step is switching to treatment with tranylcypromine (MAOI), or mirtazapine combined with extended-release venlafaxine.[14]

The MDD Clinical Practice Review Task Force released an alternative treatment model in 2016. This model proposes that psychotherapy such as cognitive behavioral therapy (CBT), interpersonal psychotherapy (IPT), or problem-solving therapy (PST) may be used as first-line treatment. At this stage monotherapy of a selective serotonin reuptake inhibitor (SSRI), serotonin-norepinephrine reuptake inhibitor (SNRI), vortioxetine, bupropion, or mirtazapine may be recommended at an adequate dose. This treatment model also suggests combined pharmacotherapy and CBT for better short-term outcomes. If the initial treatment is not tolerated or is ineffective, then there are escalations possible. Second-line psychotherapy treatments include social skills training (SST), behavioral activation (BA), or psychodynamic therapy (PT). Second-line pharmacotherapy includes tricyclic antidepressants (TCAs) and monoamine oxidase inhibitors (MAOIs), with the possibility of adjunctive medications such as atypical antipsychotics, lithium, adding a second antidepressant, buspirone, or thyroid hormone (T3).[15]

The modern controlled clinical trial developed in parallel to the advancements of antidepressant therapy. The Food and Drug Administration (FDA) was appointed as the regulatory authority over the research industry in the United States in 1938, although the significance of regulation was not recognized for decades. Following WWII, "statistically based clinical trials became a critically important part of evidence-based medicine" in the United States.[16] It was not until the thalidomide drug crisis in 1962 that the FDA exerted control over the research industry, as the crisis demonstrated the "importance of clinical trials in new drug development as well as in clinical medicine."[16]

Now, clinical trials are tightly regulated and standardized. Clinical trials are described based on their phase (I-IV), and advancement of through each phase depends on the safety and efficacy of the proposed treatment as well as FDA approval. Phase I studies include a small number of people, and the number of people included increases in each phase. In phase II studies and beyond, clinical trials are conducted by independent clinical trial centers that are contracted to recruit participants and implement the study per the sponsor's protocol. No matter how detailed the protocol is, however, problems may still arise due to the nature of research and inherent bias.

Clinical research is a fine balancing act between the accuracy and the generalizability of the data collected. While an intervention must ultimately be effective and offer benefits in a non-ideal clinical practice, research must first establish if an intervention is effective or beneficial in an ideal situation. While no ideal situation exists outside of theory, explanatory trials aim to maximize the accuracy or reliability, or what is known as the internal validity, of a study by controlling all variables. The

generalizability, or the external validity, of a study is maximized in a pragmatic study by minimizing the exclusion criteria.[17,18]

In order to maximize internal validity, some studies employ central raters (also called independent raters), who perform site-independent qualification assessments. These central raters are off-site and speak with study participants, via telephone or video conferencing, to determine participant eligibility for the study. While central raters may ideally increase internal validity, a "lack of standardization across sites and raters, poor interrater reliability, and possible scoring bias affecting the primary outcome measure contribute to a high failure rate" in psychiatric trials that employee central raters.[19,20]

This study explores if and how central rater interviews affected the screen failure rate in three clinical trials conducted at North Texas Clinical Trials.

Materials and Methods

North Texas Clinical Trials

This study was designed under the direction of my primary Scholarly Pursuit and Thesis (SPT) mentor, Brian Maynard, PhD. Dr. Maynard began North Texas Clinical Trials in 2012. He is the Director of Clinical Research and Principal Investigator at North Texas Clinical Trials. His experience in clinical drug trials extends over 15 years and includes work in both academic and industry settings. He has acted as a sub-investigator on over 100 industry-sponsored studies. Prior publications are proprietary in nature and not public knowledge. Current studies at North Texas Clinical Trials include investigations into MDD, Parkinson's disease, tardive dyskinesia, schizophrenia, bipolar disorder, and pediatric Tourette syndrome.

Study Design

We performed a retrospective comparative analysis of three phase II, double-blind, randomized, placebo-controlled clinical trials in MDD participants who were currently being treated with antidepressant therapy in order to determine the prevalence of screen failure due to off-site central rater interview.

Data Collection & Statistical Analysis

Prior to each clinical trial, health information was gathered in detail for each participant including height, weight, and past medical and surgical history. At the time of screening, vitals were assessed, and neurological and physical examinations were conducted by a qualified medical professional. Data from a 12-lead ECG, a urine sample, and fasting blood samples were collected at screening as well. Urine and

blood samples were sent off for analysis at an outside laboratory. Laboratory tests on the samples included serum chemistry (complete metabolic panel), hematology (complete blood count), thyroid function, HbA1c, lipid panel, urinalysis, and drug screen. The on-site investigator reviewed the laboratory results upon their return, documented their review, and documented any clinically relevant results.

On-site screening interviews were conducted on a pre-determined schedule at North Texas Clinical Trials. A trained professional administered baseline or screening scales per study protocol and determined whether or not the participant met the study's inclusion or exclusion criteria. If the participant was deemed eligible after their on-site interview, they would undergo a site-independent qualification assessment with a central rater in order to assess the validity of the participant's diagnosis for inclusion in the study. This assessment was conducted via telephone in all three studies. If determined to be eligible for the study by the central rater, the participant would proceed to be randomized. If at participant is deemed to be ineligible during the screening process, they are said to have failed screening or are a "screen fail." Both on-site physical documentation and electronic documentation were maintained during this process.

Data retrieval during the retrospective comparative analysis was conducted through review of de-identified physical documentation for each study. The data were verified through review of de-identified electronic records. Data collected included demographic data (age, gender, ethnicity), on-site scale scores, central rater interview scale scores, current antidepressant, and, if applicable, the reason why the participant failed screening.

Statistical analysis was conducted on site by experienced members of the North Texas Clinical Trials team. Chi-squared analysis was used to determine significance of demographic variables.

Relevant Scales

Per the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-V), the diagnosis of MDD requires ≥5 of the following symptoms, including depressed mood and/or loss of interest or pleasure, during the same two-week period. These symptoms must be a change from previous functioning and must not be clearly attributable to another medical condition.

1. Little interest or pleasure in doing things
2. Feeling down, depressed, or hopeless
3. Trouble falling or staying asleep, or sleeping too much
4. Feeling tired or having little energy

5. Poor appetite or overeating
6. Feeling bad about yourself—or that you are a failure or have let yourself or your family down
7. Trouble concentrating on things, such as reading the newspaper or watching television
8. Moving or speaking so slowly that other people could have noticed. Or the opposite—being so fidgety or restless that you have been moving around a lot more than usual
9. Thoughts that you would be better off dead or of hurting yourself in some way

In addition to ≥5 of the above depressive symptoms, the following criteria are required for diagnosis.

1. Symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.
2. Episode not attributable to physiological effects of a substance or another medical condition.
3. Episode not better explained by schizoaffective disorder, schizophrenia, schizophreniform disorder, delusional disorder, or other specified and unspecified schizophrenia spectrum and other psychotic disorders
4. No history of manic or hypomanic episode
    a. This exclusion does not apply if all of the manic-like or hypomanic-like episodes are substance induced or are attributable to the physiological effects of another medical condition.

The Structured Clinical Interview for DSM-V (SCID-5) is a semi-structured interview guide for the DSM-V diagnosis of MDD. It is administered by a clinician or clinically trained researcher who is familiar with the DSM-V classification and diagnostic criteria as well as clinical diagnostics. The SCID-5-CT is an adaptation of the SCID-5 that has been optimized for use in clinical trials that incorporate typical inclusion and exclusion criteria. The SCID-5-CT typically takes 30 to 75 minutes to administer, and the number of questions is customizable to each study's protocol.[21] This scale was administered during an on-site screening interview by a clinically trained researcher to determine the diagnosis of MDD in Study 2 and 3. A patient must be determined to have MDD by the researcher in order to meet inclusion criteria. A patient was excluded from the study if they had a history or current diagnosis of a psychotic disorder, bipolar disorder, intellectual disability, autism spectrum disorder, borderline personality disorder, or a somatoform disorder.

The Columbia-Suicide Severity Rating Scale (C-SSRS) distinguishes between suicidal ideation and suicidal behavior by measuring the severity of suicidal ideation, the intensity of suicidal ideation, suicidal behavior, and lethality of suicide attempts. The severity of suicidal ideation is graded on a scale of 1

("wish to be dead") to 5 (suicidal plan with intent). The intensity of suicidal ideation measures five items (frequency, duration, controllability, deterrents, and reason for ideation) on a scale of 1 to 5 that indicates increasing severity. The section on suicidal behavior measures actual, aborted, and interrupted suicide attempts, as well as preparatory behavior and non-suicidal self-injurious behavior. The actual lethality of the suicide attempts is graded on a scale of 0 (very minor or no physical damage) to 5 (death). If the patient scores a 0 on the actual lethality, then potential lethality is measured on a scale of 0 ("behavior not likely to result in injury") to 2 ("behavior likely to result in death despite available medical care"). The C-SSRS can be administered to measure suicidal ideation and suicidal behavior in any given timeframe, such as the time since the last appointment, the past six months, or a lifetime. Measurement of suicidal ideation and suicidal behavior over the course of a lifetime asks the patient to recall the time that they felt the most suicidal.[22,23] It is important to get a baseline score as well as determine that the patient is not a risk to themselves. The C-SSRS was administered by a clinically trained researcher during an on-site screening interview in Study 1, 2, and 3.

The Hamilton Depression Rating Scale (HDRS-17) "is one of the longest standing, most widely used measures of depression severity in research and clinical practice."[24] The scale contains 17 items pertaining to symptoms of depression experienced over the last week. The items are rated on a Likert scale of varying value and are summed to produce a total score. A score of 0-7 is considered to be within the normal range, 8-16 suggest mild depression, 17-23 moderate depression, and scores over 24 are indicative of severe depression. The maximum score on the 17-point scale is 52.[25]

1. Depressed mood
2. Feelings of guilt
3. Suicide
4. Insomnia: early in the night
5. Insomnia: middle of the night
6. Insomnia: early hours of the morning
7. Work and activities
8. Retardation
9. Agitation
10. Anxiety psychic
11. Anxiety somatic (physiological concomitants of anxiety)
    a. Gastrointestinal: dry mouth, wind, indigestion, diarrhea, cramps, belching
    b. Cardiovascular: palpitations, headaches
    c. Respiratory: hyperventilation, sighing

    d. Urinary frequency
    e. Sweating
12. Somatic symptoms gastro-intestinal
13. General somatic symptoms
14. Genital symptoms (loss of libido, menstrual disturbance, erectile dysfunction)
15. Hypochondriasis
16. Loss of weight (either according to patient or according to weekly measurements)
17. Insight

HDRS-17 was administered by an offsite central rater screening interviews in Study 2 and 3. A score less than 20 was exclusionary, as was an improvement of greater than 20% between central rater interviews.

The Montgomery-Åsberg Depression rating scale (MADRS) is designed to measure depression severity and detects changes due to antidepressant treatment. The ten items are clinician-rated on a seven-point Likert scale from 0 (normal or not present) to 6 (severe or continuous presence of symptoms), and are summed to produce a total scale score, with higher scores reflecting greater depression severity.
A score of 0-6 is considered to be within normal limits, while a score 7-19 indicates mild depression, 20-34 indicates "moderate depression," a score of 35 and greater indicates "severe depression," and a score of 60 indicates very severe depression. This scale has high inter-rater reliability, meaning that there is a high degree of agreement between independent assessments. Furthermore, there is evidence that an improvement of two points or more on the MADRS is considered clinically relevant.[26]

1. Apparent sadness
2. Reported sadness
3. Inner tension
4. Reduced sleep
5. Reduced appetite
6. Concentration difficulty
7. Lassitude
8. Inability to feel
9. Pessimistic thoughts
10. Suicidal thoughts

MADRS was administered by a clinically trained researcher during an on-site screening interview in Study 1, as well as during the offsite central rater screening interview. A score less than 23 was considered exclusionary.

The Massachusetts General Hospital Antidepressant Treatment History Questionnaire (MGH-ATRQ) is a self-reported scale used to determine treatment response and resistance in MDD. It evaluates the adequacy of duration and dose of all antidepressant medications taken by the patient during their current major depressive episode. The scale provides specific criteria for adequate dosage for the most commonly used antidepressants. The scale defines 6 weeks on an adequate dose of antidepressant medication as an adequate duration of treatment. In addition, the MGH-ATRQ assesses the degree of improvement on a scale from 0% (not improved at all) to 100% (completely improved).[27] Several criteria were exclusionary on the MGH-ATRQ during the central rater interview in Study 2 and 3. The degree of improvement was exclusionary if it was equal to or greater than 50%. A large degree of improvement in mood after starting a medication followed by a sudden decrease in mood, called tachyphylaxis, as determined by the investigator was exclusionary. Certain medications, such as benzodiazepines, were exclusionary in both studies as well.

Study	On-site scales	CR scales	Number of CR interviews
1	MADRS C-SSRS	MADRS	1
2	SCID-CT C-SSRS	HDRS-17 MGH-ATRQ	2
3	SCID-CT C-SSRS	HDRS-17 MGH-ATRQ	2

Table 1. Screening scales by study

Results

Study	Study participants	Gender	Average Age (Years)	Race	Ethnicity	Enrolled	Failed before CR interview	Failed at CR interview
1	N=12	Male: 6 Female: 6	37.9	White: 11 Other: 1	Not H/L*: 9 H/L: 1	5	5	2
2	N=15	Male: 4 Female: 11	48	White: 13 Black or African American: 1 Asian: 1	Not H/L: 15 0 H/L: 0	5	5	5
3	N=39	Male: 10 Female: 29	48.6	White: 34 Black or African American: 3 Asian:1 Other: 1	Not H/L: 32 H/L: 6 Italian: 1	18	11	10
Total	N=67	Male: 20 Female: 47	46.2	White: 59 Black or African American: 4 Asian: 2 Other: 2	Not H/L: 57 H/L: 9 Italian: 1	28	21	17

Table 2. Patient demographics by study * H/L – Hispanic or Latino

Status	Gender	Avg. Age (Years)	Race	Ethnicity	Medications
Enrolled: 28	Male: 6 Female: 22	42.1	White: 26 Asian: 1 Black, African American: 1	Not H/L: 22 H/L: 6	Fluoxetine, bupropion: 2 Venlafaxine, bupropion: 1 Citalopram, bupropion: 1 Sertraline: 4 Duloxetine: 4 Fluoxetine: 6 Venlafaxine: 3 Citalopram: 4 Escitalopram: 1 Paroxetine: 2
SF before CR: 21	Male: 10 Female: 11	50.7	White: 16 Black, African American: 3 Asian: 1 Other: 1	Not H/L: 20 H/L: 1	Duloxetine, bupropion: 2 Citalopram, lithium: 1 Citalopram: 2 Sertraline: 5 Unknown: 5 Duloxetine: 2 Venlafaxine: 1 Escitalopram: 2 Fluoxetine: 2
SF at CR: 17	Male: 6 Female: 11	42.4	White: 16 Other: 1	Not H/L: 14 H/L: 2 Italian: 1	Unknown: 3 Paroxetine: 1 Escitalopram: 5 Duloxetine: 2 Fluoxetine: 2 Sertraline: 1 Desvenlafaxine: 1 Venlafaxine: 1 Bupropion: 1

Table 3. Patient demographics by enrollment status overall

Study 1

Status	Gender	Avg. Age (Years)	Race	Ethnicity	Onsite score, CR score (MADRS)	Avg. onsite score, CR score (MADRS)	Medications
Enrolled: 5 (41.7%)	Male: 2 Female: 3	39.2	White: 5	Not H/L: 4 H/L: 1	36, 35 27, 36 38, 37 31, 36 35, 31	33.4, 35	Fluoxetine, bupropion: 1 Sertraline: 2 Duloxetine: 1 Fluoxetine: 1
SF before CR: 5 (41.7%)	Male: 4 Female: 1	37.8	White: 4 Other: 1	Not H/L: 4 H/L: 1	35, N/A N/A, N/A N/A, N/A 29, N/A 31, N/A	32.5, N/A	Citalopram: 1 Sertraline: 1 Duloxetine, bupropion: 2 Unknown: 1
SF at CR: 2 (16.6%)	Male: 0 Female: 2	35	White: 2	Not H/L: 1 H/L: 1	35, 21 36, 31	35.5, 26	Unknown: 2

Table 4. Study 1 demographics by enrollment status

Study 1 Screen Failures

5 study participants were excluded prior to central rater interview for the following reasons:

- Did not complete central rater interview (1).
- Withdrew consent before screening scales were administered (1).
- Exclusion due to abnormal lab values (3 total; 2 participants had TSH values out of normal range, 1 participant had liver function tests over two times the normal range).

2 study participants were excluded due to central rater interview for the following reasons:

- MADRS on-site was 36, failed due to MADRS score of 21 during central rater interview.
- MADRS on-site was 35, failed central rater interview due to remote history of substance abuse.

Study 2

Status	Gender	Avg. Age (Years)	Race	Ethnicity	CR scores (HDRS, HDRS)	Avg. CR scores (HDRS)	Medications
Enrolled: 5 (33.3%)	Male: 1 Female: 4	41.4	White: 4 Asian: 1	Not H/L: 5 H/L: 0	25, 29 30, 33 21, 18 24, 24 32, ?	26.4, 26	Citalopram: 1 Duloxetine: 1 Fluoxetine: 1 Venlafaxine, bupropion: 1 Venlafaxine: 1
SF before CR: 5 (33.3%)	Male: 2 Female: 3	61	White: 4 Black or African American: 1	Not H/L: 5 H/L: 0	N/A	N/A	Duloxetine: 2 Sertraline: 2 Unknown: 1
SF at CR: 5 (33.3%)	Male: 2 Female: 3	46	White: 5	Not H/L: 5 H/L: 0	31, N/A 17, N/A 11, N/A 14, N/A 14, N/A	17.4, N/A	Duloxetine: 1 Escitalopram: 2 Fluoxetine: 1 Paroxetine: 1

Table 5. Study 2 demographics by enrollment status

Study 2 Screen Failures

5 study participants were excluded prior to central rater interview for the following reasons:

- ECG abnormalities (2).
- Withdrew consent before screening scales were administered (2).
- Did not complete central rater interview (1).

5 study participants were excluded due to central rater interview for the following reasons:

- Central rater HDRS score 31, but failed MGH-ATRQ due to tachyphylaxis (1)
- Central rater HDRS score 17, but failed MGH-ATRQ due to nightly doxepin (1)
- Central rater HDRS score 17, but failed MGH-ATRQ due to 50% improvement in symptoms after initiation of SSRI (1)
- Central rater HDRS score 11 (1)
- Determined to have PTSD by central rater (1)

Study 3

Status	Gender	Avg. Age (Years)	Race	Ethnicity	CR scores (HDRS, HDRS)	Avg. CR scores (HDRS)	Medications
Enrolled: 18 (46.2%)	Male: 3 Female: 15	45.7	White: 14 Black or African American: 1	Not H/L: 10 H/L: 5	29, 32 35, 31 21, 27 45, 21 30, 34 23, - 24, - 22, - 10 w/o scores	28.6, 29	Citalopram, bupropion: 1 Fluoxetine, bupropion: 1 Citalopram: 3 Fluoxetine: 4 Duloxetine: 2 Escitalopram: 1 Sertraline: 2 Venlafaxine: 2 Paroxetine: 2
SF before CR: 11 (28.2%)	Male: 4 Female: 7	53.3	White: 8 Black or African American: 2 Asian: 1	Not H/L: 11 H/L: 0	N/A	N/A	Venlafaxine: 1 Unknown: 2 Citalopram:1 Citalopram, lithium: 1 Escitalopram: 2 Fluoxetine: 2 Sertraline: 2
SF at CR: 10 (25.6%)	Male: 3 Female: 7	46.2	White: 9 Other: 1	Not H/L: 8 H/L: 1 Italian: 1	29, N/A 18, N/A 27, 21 18, N/A 26, 20 19, N/A 19, ? 28, 26 39, 29 17, N/A	24, 24	Escitalopram: 3 Fluoxetine: 1 Sertraline: 1 Duloxetine: 1 Unknown: 1 Desvenlafaxine: 1 Venlafaxine: 1 Bupropion: 1

Table 6. Study 3 demographics by enrollment status

Study 3 Screen Failures

11 study participants were excluded prior to central rater interview for the following reasons:

- History or current diagnosis of a psychotic disorder, bipolar disorder (1)
- Enrolled in PTSD study in last 12 months (1)
- Exclusion due to abnormal lab values (5 total; 1 had elevated TSH, 3 had elevated HbA1c values, and 1 had proteinuria with abnormal GFR and creatinine)
- Active alcohol use disorder (1)
- Withdrew consent before screening scales were administered (2).
- Opinion of on-site investigator (MDD severity) (1)

10 study participants were excluded due to central rater interview for the following reasons:

- "Any condition for which, in the opinion of the investigator, participation would not be in best interest of the participant or that could prevent, limit, or confound the protocol-specified assessments." (1)
- Central rater HDRS score 18 (2)
- Central rater HDRS score 19 (2)
- Central rater HDRS score 17 (1)
- HDRS score improved >20% from the first to second independent HDRS rating (4)

Age

Outcome	0	1	Total
0	10	7	17
1	16	12	28
Total	26	19	45

Table 7. Contingency table: age. Key: 0= 40 and below, 1= 41 and above.

	Value	df	p
X^2	.012	1	0.912
N	45		

Table 8. Chi-squared test: age

Gender

Outcome	0	1	Total
0	12	5	17
1	22	5	28
Total	34	11	45

Table 9. Contingency table: gender. Key: 0= Female, 1= Male.

	Value	df	p
X^2	0.365	1	0.546
N	45		

Table 10. Chi-squared test: gender

Class of antidepressant therapy (ADT)

Outcome	0	1	2	Total
0	12	4	1	17
1	19	8	1	28
Total	31	12	2	45

Table 11. Contingency table: class of ADT. Key: 0= SSRI, 1=SNRI, 2= Other

	Value	df	p
X^2	0.239	2	0.887
N	45		

Table 12. Chi-squared test: class of ADT

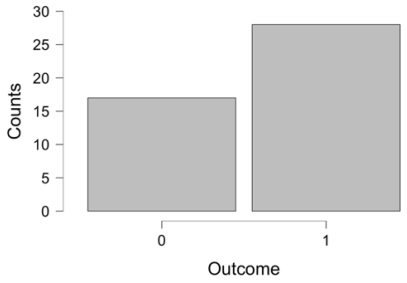

Fig. 1. Distribution plot: outcome

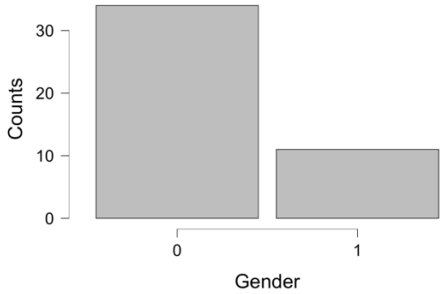

Fig. 2. Distribution plot: gender

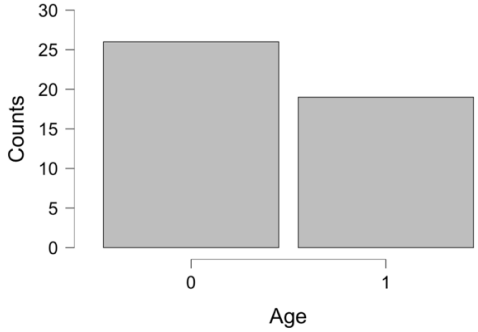

Fig. 3. Distribution plot: age

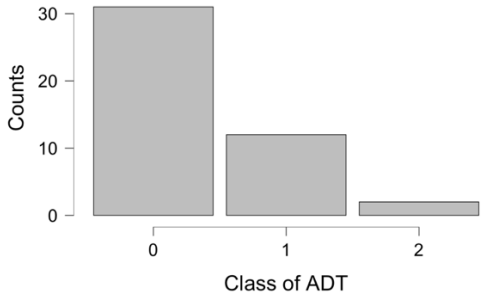

Fig. 4. Distribution plot: class of ADT

Discussion & Innovation

In Study 1 (N=12), 58.3% of participants failed at screening, compared to 66.6% of participants in Study 2 (N=15), and 58.3% of participants in Study 3 (N=39). These findings are in line with data from the Tufts Center for the Study of Drug Development (Tufts CSDD) that indicated that the average screen failure rate for CNS clinical trials (in which category major depressive disorder is classified) was 57% in 2019.[28] The demographic data, specifically age, gender, and antidepressant therapy, were statistically insignificant ($p = 0.912$, $p = 0.546$, $p = 0.887$) in relation to screen failure rate.

In Study 1, 41.7% of participants failed screening prior to central rater call, whereas 33.3% and 28.2% of participants failed screening prior to the central rater call in Study 2 and Study 3 respectively. Of the 21 participants that failed prior to central rater call, 10 participants were excluded due to abnormal lab or ECG values and 7 participants withdrew consent or did not complete the central rater interview.

In Study 1, 16.2% of participants failed screening at the central rater call, whereas 33.3% and 25.6% failed screening at the central rater call in Study 2 and Study 3 respectively. Of the 17 participants that failed at central rater call, 14 participants were excluded due to MDD severity measured by scale scores and 3 were excluded by central rater due to other exclusion or inclusion criteria.

In Study 2, 60% of those who failed screening at the central rater interview did so due to the MGH-ATRQ scale. This scale was not administered during the onsite interview, which may bias the results slightly. However, the MGH-ATRQ scale was also not administered during the onsite interview in Study 3 and no study participants failed screening at the central rater call due to the results of the MGH-ATRQ scale. Instead, 90% of those who failed screening at the central rater call in Study 3 did so due to their absolute HDRS-17 score or a change of 20% or more in their score.

A trend emerged while comparing the causes of screen failure for participants who failed screening prior to central rater interview to those who failed screening at central rater interview. Those who failed screening prior to central rater interview were most likely to fail due to inclusion and exclusion criteria other than MDD severity. Those who failed screening at central rater interview were most likely to fail due to MDD severity or treatment response measured by the MGH-ATRQ scale.

There were several limitations to this study. The overall sample size (N=67) was smaller than predicted when initially approaching the data, which may limit the generalizability of this study. There were protocol changes to Study 3 partway through enrollment due to the high number of screen failures at the central rater interview at all participating sites. During the beginning of Study 3, there were few communicated records of central rater scoring, making it difficult to evaluate the internal validity of their scoring.

As mentioned above, Study 1 is unique in that the same screening scale is used at both the onsite interview and the offsite central rater interview, as well as only having one central rater interview. In both

Study 2 and Study 3, the scale administered during the onsite interview differs from the scale administered during the central rater interview. This makes it difficult to compare scores between the onsite and central rater interview. Given these challenges, it can be difficult to make direct comparisons between scores from two different scales administered by two different raters. Efforts can be made to ensure the scales are equivalent and administered in similar ways.

While the use of off-site central raters has been examined through a technical lens, it may also be important to understand their impact on study participant perception and communication. There are several reasons as to why this may bias the self-reporting of patients or interviewer assessments.

Off-site central rater interviews were conducted via telephone in all three studies examined in this retrospective analysis. A review of the literature revealed that use of telephone interviews or video conferencing in the current technological climate has not been assessed extensively, as a majority of the articles were published several decades ago. While several studies suggest objective agreeability between telephone interviews and in-person interviews when using certain scales during the evaluation of psychiatric conditions, others have demonstrated that there is variability in disclosure depending on the topic of discussion.[29,30,31] From a patient's perspective, technology may offer freedom to speak about their experience with someone who is physically removed from their life. But for others, this may be a barrier to trust that would otherwise be built in-person.

Remote interviews, especially those conducted via the telephone, also may not capture the patient's full gestalt as readily. Nonverbal behavior in patients with major depressive disorder, while dependent on the interpretation of the interviewer, has been shown to differ from nonclinical subjects.[32,33] This element of the initial clinical assessment is missing during a telephone interview.

Given these potential effects, future research could establish how to leverage study participant perception and technological communication to decrease bias and increase internal validity when integrating site-independent qualification assessments into research protocol, such as in decentralized clinical trials.

Compliance Plan

This project was conducted under North Texas Clinical Trials' central IRB.

This research study was conducted at North Texas Clinical Trials under the supervision of my primary SPT mentor, Dr. Brian Maynard. All patients were evaluated at North Texas Clinical Trials, which kept detailed medical records of each patient, including baseline vitals, a baseline ECG, and blood tests, along with all follow up medical testing. All physical records were kept at North Texas Clinical Trials, but I, along with other members of the team, were able to access de-identified data on our personal laptops.

References

1. Hasin DS, Sarvet AL, Meyers JL, et al. Epidemiology of Adult DSM-5 Major Depressive Disorder and Its Specifiers in the United States. JAMA Psychiatry. 2018;75(4):336-346. doi:10.1001/jamapsychiatry.2017.4602.
2. Fergusson DM, McLeod GF, Horwood LJ, et al. Life satisfaction and mental health problems (18 to 35 years). Psychol Med 2015; 45:2427. doi: 10.1017/S0033291715000422.
3. Ferrari AJ, Charlson FJ, Norman RE, et al. Burden of depressive disorders by country, sex, age, and year: findings from the global burden of disease study 2010. PLoS Med 2013. doi: 10.1371/journal.pmed.1001547
4. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5), American Psychiatric Association, Arlington 2013.
5. DSM-5 Criteria: Major Depressive Disorder. Medicaid Mental Health. USF, pp.24-32. Published 2017. https://medicaidmentalhealth.fmhi.usf.edu/_assets/file/Guidelines/2017-2018%20Treatment%20of%20Adult%20Major%20Depressive%20Disorder.pdf. Accessed 7 June 2020.
6. Judd LL, Schettler PJ, Coryell W, et al. Overt irritability/anger in unipolar major depressive episodes: past and current characteristics and implications for long-term course. JAMA Psychiatry 2013; 70:1171. doi: 10.1001/jamapsychiatry.2013.1957
7. Rock PL, Roiser JP, Riedel WJ, Blackwell AD. Cognitive impairment in depression: a systematic review and meta-analysis. Psychol Med 2014; 44:2029. doi: 10.1017/S0033291713002535
8. Bora E, Berk M. Theory of mind in major depressive disorder: A meta-analysis. J Affect Disord 2016; 191:49. doi: 10.1016/j.jad.2015.11.023
9. Trivedi MH, Greer TL. Cognitive dysfunction in unipolar depression: implications for treatment. J Affect Disord 2014; 152-154:19. doi: 10.1016/j.jad.2013.09.012
10. Maurer DM. Screening for Depression. Texas Am Fam Physician. 2012 Jan 15;85(2):139-144.
11. Hillhouse TM, Porter JH. A brief history of the development of antidepressant drugs: From monoamines to glutamate. Exp Clin Psychopharmacol 2015;23(1):1-21. doi: 10.1037/a0038550
12. López-Muñoz F, Álamo C. History of the Discovery of Antidepressant Drugs. Melatonin, Neuroprotective Agents and Antidepressant Therapy. Springer 2016.
13. Kishimoto T, Chawla JM, Hagi K, et al. Single-dose infusion ketamine and non-ketamine N-methyl-d-aspartate receptor antagonists for unipolar and bipolar depression: a meta-analysis of efficacy, safety and time trajectories. Psychol Med 2016; 46:1459. doi: 10.1017/S0033291716000064
14. Sequenced Treatment Alternatives to Relieve Depression STAR*D Study. NIMH. 2006.
15. Halverson J, Beevers C, et al. Clinical Practice Review for Major Depressive Disorder. ADAA. 2016. https://adaa.org/resources-professionals/practice-guidelines-mdd
16. Junod SW, Beaver WT. FDA and Clinical Drug Trials: A short history. FDA. https://www.fda.gov/media/110437/download. Published 2008.
17. Godwin M, Ruhland L, Casson I, et al. Pragmatic controlled clinical trials in primary care: the struggle between external and internal validity. BMC Med Res Methodol. 2003;3:28. Published 2003 Dec 22. doi:10.1186/1471-2288-3-28
18. Steckler A, McLeroy KR. The importance of external validity. Am J Public Health. 2008;98(1):9-10. doi:10.2105/AJPH.2007.126847

19. Williams JB, Kobak KA, Giller E, Reasner DS, Curry L, Detke MJ. Comparison of Site-Based Versus Central Ratings in a Study of Generalized Anxiety Disorder. J Clin Psychopharmacol. 2015;35(6):654-660. doi:10.1097/JCP.0000000000000422Chen
20. Carleton RN, Tibodeau MA, et al. The Center for Epidemiologic Studies Depression Scale: A Review with a Theoretical and Empirical Examination of Item Content and Factor Structure. PLoS One 2013; 8(3): e58067. doi: 10.1371/journal.pone.0058067
21. The Structured Clinical Interview for DSM-5®. APPI. https://www.appi.org/products/structured-clinical-interview-for-dsm-5-scid-5.
22. Posner K, Brent D, et al. Columbia-Suicide Severity Rating Scale. New York State Psychiatric Institute. Research Foundation for Mental Hygiene. 2010. https://cssrs.columbia.edu/wp-content/uploads/C-SSRS_Pediatric-SLC_11.14.16.pdf
23. Columbia-Suicide Severity Rating Scale (C-SSRS). National Suicide Prevention Lifeline. https://pcl.psychiatry.uw.edu/wp-content/uploads/2021/12/C-SSRS.pdf
24. Rohan KJ, Rough JN, Evans M, et al. A protocol for the Hamilton Rating Scale for Depression: Item scoring rules, Rater training, and outcome accuracy with data on its application in a clinical trial. J Affect Disord. 2016;200:111-118. doi:10.1016/j.jad.2016.01.051
25. Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry 1960; 23:56–62 https://dcf.psychiatry.ufl.edu/files/2011/05/HAMILTON-DEPRESSION.pdf
26. Aripiprazole (Abilify): Depression, Major Depressive Disorder (MDD) [Internet]. Ottawa (ON): Canadian Agency for Drugs and Technologies in Health; 2016 Nov. APPENDIX 5, VALIDITY OF OUTCOME MEASURES. Available from: https://www.ncbi.nlm.nih.gov/books/NBK409740/
27. Chandler GM, Iosifescu DV, Pollack MH, Targum SD, Fava M. RESEARCH: Validation of the Massachusetts General Hospital Antidepressant Treatment History Questionnaire (ATRQ). CNS Neurosci Ther. 2010;16(5):322-325. doi:10.1111/j.1755-5949.2009.00102.x
28. Getz, K. Can Recruitment and Retention Get Any Worse? Applied Clinical Trials. 2019;28(12). https://www.appliedclinicaltrialsonline.com/view/can-recruitment-and-retention-get-any-worse
29. Kobak KA, Leuchter A, DeBrota D, et al. Site versus centralized raters in a clinical depression trial: impact on patient selection and placebo response. J Clin Psychopharmacol. 2010;30(2):193-197. doi:10.1097/JCP.0b013e3181d20912
30. Hajebi A, Motevalian A, Amin-Esmaeili M, et al. Telephone versus face-to-face administration of the Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, for diagnosis of psychotic disorders. *Compr Psychiatry*. 2012;53(5):579-583. doi:10.1016/j.comppsych.2011.06.001
31. Shuy RW. In-Person Versus Telephone Interviewing. In Holstein JA, Inside Interviewing: New Lenses, New Concerns. Sage Publications; 2003: 175-192.
32. Fiquer JT, Moreno RA, Brunoni AR, Barros VB, Fernandes F, Gorenstein C. What is the nonverbal communication of depression? Assessing expressive differences between depressive patients and healthy volunteers during clinical interviews. *J Affect Disord*. 2018;238:636-644. doi:10.1016/j.jad.2018.05.071
33. Annen S, Roser P, Brüne M. Nonverbal behavior during clinical interviews: similarities and dissimilarities among schizophrenia, mania, and depression. *J Nerv Ment Dis*. 2012;200(1):26-32. doi:10.1097/NMD.0b013e31823e653b