IRB Protocol Submission

A Comparison of Cognitive and Dynamic Therapy for MDD in Community Settings

Principal Investigator: Mary Beth Gibbons, Ph.D.
Associate Professor of Psychology in Psychiatry
Department of Psychiatry
University of Pennsylvania
3535 Market St., Suite 649
Philadelphia, PA 19104-4283

1. Study Sponsorship
This study is funded by the Agency for Healthcare Research and Quality

2. Introduction and Purpose
A. SPECIFIC AIMS
The overall goal of this study is to conduct a trial that tests the hypothesis that supportive-expressive therapy, a type of psychodynamic psychotherapy, is not inferior to cognitive therapy when used for the treatment of depression in community mental health consumers. Supportive-expressive dynamic psychotherapy focuses on changes in self-understanding of interpersonal patterns as one of the main curative factors of psychotherapy. Cognitive therapy targets behavioral activation and disconfirmation of specific negative expectations and thoughts.

3. Objectives
The overall goal of this proposal is to conduct a randomized, comparative, non-inferiority clinical trial that tests the hypothesis that a widely used form of manualized dynamic psychotherapy (supportive expressive psychodynamic therapy) is not inferior to cognitive therapy when implemented in community mental health settings for the treatment of major depressive disorder. The specific aims are to conduct a randomized non-inferiority trial to compare supportive-expressive psychodynamic therapy and cognitive therapy for patients with major depressive disorder and assess the comparative effectiveness of supportive-expressive psychodynamic therapy and cognitive therapy on secondary measures of symptoms, patient functioning, and quality of life.

Primary Outcome Variable
The primary outcome measure is symptomatic level of depression as measured by the Hamilton Rating Scale for Depression (HAM-D). This measure will be completed at baseline as well as at months 1, 2, 4, and 5.

Secondary Outcome Variable
Secondary outcomes will include functional, symptomatic, and quality of life measures. For the secondary analyses we will use the BASIS-24 total score, the BDI-II score, the Physical Component Summary and the Mental Component Summary of the SF-36 as well as the total score from the QOLI. These measures will be completed at baseline as well as at months 1, 2, 4, and 5, with the exception of the BASIS-24 and BDI-II, which are also completed weekly.

4. Background
There is substantial evidence supporting the efficacy of cognitive therapy (Beck, Rush, Shaw, & Emery, 1979) in the treatment of MDD. Cognitive therapy has been the most widely studied psychotherapeutic treatment for MDD. The strong evidence in support of cognitive therapy for the treatment of MDD has led to a widespread effort to disseminate this treatment package to community settings. However, to date, there has not been a single empirical investigation demonstrating that cognitive therapy is superior to the dynamically oriented psychotherapy for MDD that is most commonly practiced in the community. Dynamically oriented psychotherapies have been and remain widely practiced in the United States and continental Europe. Whereas researchers of cognitive and behavioral therapies have traditionally been interested in obtaining objective evidence from large samples for the efficacy of their treatments, researchers of psychoanalytic and psychodynamic approaches have instead focused on process studies and studies of heterogeneous patient samples. Whatever the reason for the unfortunate lack of empirical data on the efficacy of dynamic therapy, such evidence is critical to daily clinical decision making and contemporary public health policy, given that this treatment is so widely practiced. Only one investigation specifically compared cognitive therapy to a dynamically oriented psychotherapy for the treatment of MDD. Shapiro et al. (1994) randomized 117 patients with MDD to either 8 or 16 sessions of manualized psychodynamic-interpersonal psychotherapy (PI) or cognitive behavioral psychotherapy (CBT). The authors concluded there is little evidence that CBT was more effective than PI across the outcome battery and across severity levels. There were no significant differences between treatment groups on measures of general symptoms, interpersonal problems, or social functioning. However, there was a significant difference between treatments on end of treatment scores on the BDI (p.05), favoring CBT. In the present case, a non-inferiority study is proposed because it addresses the question that is most relevant and meaningful to clinical treatment providers/agencies. If psychodynamic therapy is not inferior to cognitive therapy, the cost of re-training psychotherapists in cognitive therapy may not be warranted. A superiority trial that yields the result of no significant difference would be inadequate to inform a policy decision. Moreover, as mentioned, if cognitive therapy is truly superior, the current trial will also reveal that potential outcome.

5. Characteristics of the Study Population

Target population:
Consumers: Community mental health consumers age 18-65 with a diagnosis of major depressive disorder or depression NOS (approximately 290 consumers). Therapists: Clinicians at Northwestern Human Services who have agreed to participate in the study (approximately 20 therapists).

Inclusion criteria:
1. Score of 11 or above on the QIDS
2. Diagnosis by the intake clinician of major depressive disorder or depression NOS
3. Confirmed diagnosis of major depressive disorder based on the SCID-I
4. Able to read at least the 4th grade level
5. 18 to 65 years of age
6. Willingness to be randomized and participate in research

Exclusion criteria:
1. Current psychotic disorder that prohibits participation in outpatient services
2. Acute medical problem requiring immediate inpatient treatment
3. Current substance abuse or dependence requiring immediate referral to substance abuse program
4. Significant suicidal risk/ideation requiring immediate referral
Subject Recruitment and Screening:
As part of the IP-RISP to study interventions for depression in the community, we have developed a recruitment procedure in the community mental health system designed to identify potential candidates for research without placing additional burden on the community intake clinician. The current protocol will use the recruitment method already in place. In the community mental health system, the intake procedure is designed to provide quick diagnoses that can be used to make the best treatment recommendation possible; the goal is to get patients into treatment as soon as possible with minimal effort placed on fine tuning differential diagnoses. In our experience working in the community mental system, alternative diagnoses, such as depression NOS, are often provided for clients seeking treatment who would meet the diagnostic criteria for major depressive disorder. Our recruitment procedure is designed to build on the diagnostic system already in place in the community and also to include minimal additional diagnostic assessments required in research to clearly specify the sample. We cast a wide net by evaluating all clients diagnosed by the community intake clinician as either major depressive disorder or depression NOS but select only clients who meet criteria for major depressive disorder. This procedure makes use of the diagnosis provided by the community clinician without placing additional burdens on the community clinician to provide diagnosis at a level of specificity required by the research. By using research staff to conduct the minimal additional assessment we need to specify the diagnosis, we minimize the burden on community clinicians and patients while still selecting a well-specified sample that is representative of those diagnosed with major depressive disorder in the community. We have implemented a brief depressive symptom measure to identify patients with moderate to severe depressive symptoms who could be potential candidates for our research program. The Quick Inventory for Depressive Symptomatology (QIDS; Rush et al., 2003) is a reliable and valid 16-item, self-report scale designed to identify moderate to severe depression. Research has demonstrated that a score greater than or equal to 11 on the QIDS is comparable to a score of 14 or above on the Hamilton Rating Scale for Depression (HAM-D; Hamilton, 1960). NHS currently includes the QIDS as part of the intake assessment for all new patients. The intake clinician at NHS simply examines the QIDS completed by the patient and determines whether the patient is interested in hearing more about the research program. Any patient who scores 11 or above on the QIDS and indicates that he/she is interested in hearing more about the research program will be identified by the intake clinician as potentially eligible for the research program on major depressive disorder in the community. The intake clinician will complete the intake evaluation according to normal clinic procedures including provision of a clinical diagnosis. At the end of the intake appointment, instead of providing the patient with a therapist name and appointment time, the intake clinician will inform our research staff of the patient’s name, phone number, and preferred time of contact. A member of our research staff will then contact the patient, describe the possible research studies available, screen the patient for eligibility, and schedule an in-person baseline evaluation at the patients convenience. The baseline evaluation, conducted by our research staff at the agency, includes the HAM-D and the Structured Clinical Interview for the DSM-IV (APA, 1994) Axis I disorders (SCID-I; First et al., 1997). Any patient who meets criteria for the protocol will then be set up for their first therapy appointment with a therapist participating in the protocol. Any patient who does not meet criteria for inclusion in the research protocol will be immediately returned to the original agency intake clinician for assignment to a community clinician.

6. Study Design, Methods, and Procedures
The research team plans to conduct a randomized, comparative, non-inferiority clinical trial that tests the hypothesis that a widely used form of manualized dynamic psychotherapy (supportive expressive psychodynamic therapy) is not inferior to cognitive therapy when implemented in community mental health settings for the treatment of major depressive disorder. We will
implement a training phase prior to randomization to train therapists to competence in both SE and CT for participation in this trial. We will also implement a training program to run concurrently with the randomized phase in order to retain the appropriate number of therapists throughout the study. We will randomize patients seeking services at two community mental health centers to either receive 16 sessions of cognitive therapy or psychodynamic therapy. The cognitive therapy and psychodynamic therapy conditions will be active at the same time, and consumers will not cross over from one condition to the other. Measures will be completed at sessions and at baseline, month 1, month 2, month 4, and month 5 assessments. This study will take place over five years. A consumer’s participation in the study will last no longer than five months (including all therapy sessions and assessments). The estimated length of a therapist's participation could last up to the five years of the study. The project start date is September 30, 2009.

Therapist Recruitment
Therapists will consent to participation. Their sessions with study consumers will be audio recorded. They will be compensated for training, supervision, and treating consumers. Group supervision will take place twice monthly at a location deemed mutually acceptable to therapists and supervisors. Therapists will also participate in up to four individual supervision sessions per training case.

Subject Recruitment
Consumers will complete the QIDS at intake, as per usual NHS intake procedures. Following intake, consumers scoring 11 or higher on the QIDS and answering "yes" to the research interest questions will discuss the study with a Research Assistant. During this discussion the Research Assistant will conduct a phone screen to determine preliminary eligibility for participation in the study.

During the baseline assessment, consumers will read the baseline consent form, ask questions, and sign the form if they wish to participate in the baseline assessment. The baseline assessment will consist of the SCID-I, BASIS-24, BDI-II, HAM-D, QOLI, SF-36, demographic questionnaire, treatment expectations form, IIP, TLEQ, AUDIT, DAS, WOR, PDST, and SUIP-R.

Subject Participation
If the consumer qualifies for the study following the baseline assessment, the consumer will read the study consent form and HIPAA form. The consumer will ask the Research Assistant any questions he or she may have about these forms and the study. If the consumer wishes to participate in the study, the consumer will sign the forms. Consumers will be given copies of all forms for their records. Following consent into the study, consumers will be randomized to SE or CT. Consumers will attend 16 weekly therapy sessions. Consumers will complete the BASIS-24 and BDI-II at each weekly session. Consumers will attend assessments at 1 month, 2 months, 4 months, and 5 months into treatment. Consumers will also complete a treatment credibility form after their second session of therapy, and the WAI after their second, fourth, sixth, and 8th sessions. Consumers will be contacted to schedule assessments at their convenience. We will try to schedule them for before or after a therapy session, but if this is not convenient for the consumer, the assessment will be scheduled at another date/time that is convenient. A reminder call will be made if this scheduling occurs more than a day or two before the assessment. When scheduling assessments, the general script is as follows: "Hello, this is ___ from the University of Pennsylvania. You are due for your ___ month assessment soon, and I was hoping we could schedule it for the same day as your session next week if that's convenient for you. Is that okay?" For assessment reminder calls, the general script is as follows: "Hello, this is ___ from
the University of Pennsylvania. This is a reminder about your meeting on ___ at __o'clock with ___.

Chart Review Follow-Up for Enrolled Consumers
Upon completion of the consumer's participation in the study, the Research Assistant will obtain the consumer's medication information from the consumer's patient chart. Consumers who do not wish to participate at the time of consenting or discontinue participation in the study will still be eligible for treatment at the clinic. The Research Assistant will contact the Intake Coordinator to ensure the consumer is assigned a therapist.

Method for Assigning Subjects to Groups
Randomization will be done separately for the active treatment patients and the training cases. Within each of the two community mental health centers, active treatment patients will be randomized in fourteen blocks of eight (n=112) using a computer-generated random number program, with the constraint that each block must contain four assignments to SE and four to CT. This method will ensure equal group sizes among the 112 randomized patients within a site. Also, this guarantees balance between assignments for every eight randomized patient in the event of early stoppage or lower than expected accrual rate. The remaining 3 active treatment patients per site will be randomized as follows: 2 SE patients and 1 CT patient within one site and 2 CT patients and 1 SE patient in the other site. Site with the 1 additional CT patient will be chosen by a coin. Order of the treatment assignment of these last three patients will be based on a computer-generated random number program. This will result in 57 CT and 58 SE patients within the first community center and 58 CT and 57 SE patients within the other community center. This represents the full target sample proposed for the study. Patient assignment to therapist will be done sequentially, where the ordering of therapist will be based on their numerical computer generated unique identifier. Upon therapist attrition, new therapist will replace the therapist who left the study in the sequence. Randomization of training cases will mirror the active treatment in that the assignment is based on a computer-generated random number program, but unlike the active treatment assignment, randomization of treatment assignments will be done in blocks of two. Therefore, with each block, there is one SE and one CT patient. Performing the randomization in block of two extends the randomization beyond just an alternating sequence. Patient assignment to therapist will mirror as described above with the exception that when a therapist has achieve his/her quota no further assignment of training cases will be assigned to that therapist. Therapists will be randomized to CT or SE based on a similar procedure, though modified due to the fewer number of therapists.

Data Analysis
Preliminary data analyses will include descriptive statistics and exploratory graphing such as frequencies, means, standard deviations, box and whisker plots, stem and leaf diagrams, and scatter plots to assess the normality of the data in terms of the presence of skew and/or outliers for both the outcomes and adherence scores. If necessary, the continuous outcome data will be transformed by using an appropriate transformation such as the log transformation for skewed, long-tailed data. Hypothesis 1: Supportive-expressive therapy will be not inferior to cognitive therapy on change from baseline to endpoint of acute treatment in the HAM-D total score. This hypothesis is formulated in terms of a noninferiority test, which is a one-sided equivalence test. However, we are also interested in testing the possibility that cognitive therapy is superior to psychodynamic therapy. Hirotsu (2007) outlines a unifying approach to address equivalence and superiority tests. The unifying approach considers performing a closed testing procedure, which will first consider the test of equivalence followed by a subsequent test for treatment superiority. But in this multiple decision process, a price must be paid to account for the two
decisions; therefore, the alpha-level for this process is set at 0.025 (=0.05/2). Whereas the purpose of a traditional hypothesis test is to determine whether two groups differ from one another, equivalence testing is used to determine whether the two groups are sufficiently near each other to be considered equivalent. For Non-inferiority testing, the purpose of the test is to determine if the difference between groups is at most some negligible amount. Currently the standard for assessing equivalence is Schuirmann’s (1987) two one-sided t-tests, which is computationally equivalent to a classical 95% Confidence Interval. The Schuirmann’s interval hypothesis is as follows: H0: The two groups are not equivalent vs. H1: The two groups are equivalent H0: uA - uB ≤ L or uA - uB ≥ U vs. H1: L uA-uB U. For non-inferiority, the hypothesis is as follows: H0: uA ≥ uB vs. H1: L uA - uB U Where uA and uB are the respective test statistics for the two groups; L and U correspond to the lower and upper bound, respectively, for our non-inferiority margin in assessing equivalence and L is the negligible difference in assessing non-inferiority. From a testing perspective, the goal is to demonstrate statistically that an observed difference between the two groups is too large to have come from a distribution with mean L (Lower Bound Test) and too small to have come from a distribution with mean U (Upper Bound Test). For non-inferiority, the goal is to demonstrate statistically that an observed difference between the two groups is too large to have come from a distribution with mean L (Lower Bound Test). So the Lower Bound Test seeks to reject the null hypothesis asserting that the difference between the two means is less than or equal to L, whereas, the Upper Bound Test seeks to reject the null hypothesis asserting the difference is greater than or equal to U. While two tests are conducted, no correction for multiple comparisons is needed because the two tests are completely dependent. This approach will be done for the primary dependent variable: change from baseline to endpoint in the 17-item HAM-D total score. The non-inferiority margin will be established to be smaller than any clinically relevant change and will be based on the recommendation for this provided by Montgomery (1994) and previously implemented in a non-inferiority study of treatments for major depressive disorder (Szegedi et al., 2005): a non-inferiority margin of 2.5 HAM-D total scores points (this is the difference in change from baseline to endpoint between the two treatment groups). Standard deviation of change scores on the HAM-D are not typically reported in published studies, but we were able to calculate the relevant standard deviation in two studies of cognitive therapy for major depressive disorder (Dr. Gallop, who is the statistician for the current proposed study, was also the statistician for these two cognitive therapy trials). Standard deviations ranged from 6.4 (from Dimidjian et al., 2006) to 6.8 (DeRubeis et al., 2005) in these studies. However, one published study of medication for major depressive disorder found a standard deviation of 8.5 (Szegedi et al., 2005). Although the studies with cognitive therapy groups (Dimidjian et al. 2006; DeRubeis et al., 2005) had lower standard deviations, these studies were not done in community settings where variability would be expected to be higher. Thus, to be conservative, we use the higher figure (8.5) from the Szegedi et al. (2005) study. Based on the 8.5 standard deviation of HAM-D total score change scores, the 2.5 HAM-D points corresponds to a Cohens d effect size of 0.29. In the current context, statistical power is the probability to classify SE as non-inferior to CT when in fact it is true. The minimum total sample size of completers needed to guarantee a power of (1 - ) between two treatments is as follows: where: corresponds to quantiles on the Student t-distribution with its respective degrees of freedom. ²is the estimated variability in the outcome, is the alpha level for the test (set to 0.05, corresponding to Schuirmann’s 2 one-sided t-tests, 95% confidence intervals) is the beta level for the test (usually equal to 0.20, corresponding to 80% power) is the bound of the non-inferiority region, and ( /) is the effect size corresponding to the noninferiority margin. Under these constraints, the total sample size needed is 204.45, which corresponds to about 103 subjects per group in the analysis. Our preliminary study indicates that we will have 10% of patients with no post-baseline data; therefore, we would need to randomize 115 subjects per group. We expect that some patients will fail to complete the study; therefore, our intention-to-treat (ITT) analysis must accommodate this. We will implement
pattern mixture models to assess the randomness of the completion process. If the process is random, we will implement the modified two one-sided test approach for equivalence discussed by Lee, Kim, and Park (2005), who discussed a maximum likelihood statistical procedure for analyzing equivalence trials with missing observations. Secondary analyses will also be conducted using the mixed effects model to test for slope differences between treatment groups in the HAM-D. The mixed effects modeling is used to account for the clustering structure of the data (i.e., repeated assessments within an individual) and is implemented with the SAS Mixed Procedure of the SAS 9.1.3 software. This approach provides more information and, therefore, more power compared to cross-sectional analyses which focus on the analysis of one summary index or time-until event methods. This mixed effects modeling will allow us the flexibility to categorize the on average behavior and the trajectories or growth curves corresponding to the average behavior. This mixed effects framework is robust with respect to dropout and missing data, unless the dropout mechanism or cause of missing data is informative. For the HAM-D monthly measures, the number of potential post-baseline measures is limited to at most three with the HAM-D. We will implement a mixed-model analysis of variance (MMANOVA). The MMANOVA does not assume any specific profile/pattern between the outcome variable and time, but rather focuses on the average separation between groups across the treatment period. We will use baseline assessment as a covariate, and assess the average outcome between groups across the three post-baseline monthly assessments. The MMANOVA is similar to a repeated measures analysis of variance model, but offers flexibility to deal with missing data and modeling the variance-covariance matrix of the outcome (Schwarz, 1993). The potential impact of attrition on any treatment group effect is an important consideration. We will use pattern-mixture models to assess if there is bias due to drop out or missing data. As described by Hedeker and Gibbons (1997), these mixed models allow us to assess whether important estimates (e.g., average outcome for groups) are dependent on missing data patterns, and provide overall estimates of effects by averaging over the various missing-data patterns. We will also apply mixed-effect hybrid models (MEHM) to assess bias due to drop out. As described by Yuan and Little (in press), MEHM models the joint distribution of the outcome process and dropout process differently than the pattern mixture model. For MEHM, the joint distribution is factorized into the marginal distribution of random effects, the dropout process conditional on random effects, and the outcome process conditional on dropout patterns and random effects. All modeling structures allow for important covariates. We will include any baseline measures which differ between groups. The impact of gender and minority status on any treatment effects will also be explored in secondary analyses. Hypothesis 2: Supportive expressive therapy will be not inferior to cognitive therapy on change from baseline to endpoint of acute treatment on secondary measures of symptoms, patient functioning, and quality of life. Analyses of the secondary outcome measures will be conducted as non-inferiority analyses as described above for the primary outcome measure. For the secondary analyses we will use the BASIS-24 total score, the Physical Component Summary and the Mental Component Summary of the SF-36 as well as the total score from the QOLI. Because there is no consensual definition of a minimally clinically significant difference on these other measures, we will set the non-inferiority margin to be the same effect size ($d = 0.29$) used for the HAM-D. We will also conduct analyses using the mixed effects model to test for slope differences between treatment groups in the HAM-D on each of the secondary outcome measures, with only the number of assessments varying depending on the measure.

7. Study Instruments
Quick Inventory of Depressive Symptomatology- Self Report (QIDS; Rush et. al. 2003). The QIDS is a 16 item self-report measure designed to assess the severity of depressive symptoms using the criterion symptoms designated by the DSM-IV. The QIDS demonstrated good internal
consistency (Cronbachs alpha = .86) in patients with chronic major depression (Rush, 2003). In the same sample, total scores on the QIDS were highly correlated (r=.81) with the 17-item Hamilton Rating Scale for Depression (HAM-D; Rush, 2003).

The Structured Clinical Interview for DSM-IV (SCID; First et al., 1997). The full SCID-I for DSM-IV will be used to confirm the presence of a diagnosis of major depressive disorder and collect additional data on any comorbid diagnoses.

The Hamilton Depression Inventory (HAM-D; Hamilton, 1960). The HAM-D is a widely used inventory for evaluating the severity of common symptoms of depression. The 24-item version of the HAM-D will be completed by applying the Structured Interview Guide to enhance reliability (Williams, 1988). Using the structured interview guide, Williams reported good interjudge reliability for a test-retest assessment of the 17-item score (I = .81). A copy of the HAM-D is provided in the appendix.

The BASIS-24. All patients will complete the BASIS-24 (Eisen et al., 2004) before each outpatient session as part of routine clinical practice. The BASIS-24 is a 24-item self-report inventory designed to measure mental health status from the consumers point of view. The items cover six domains including: depression/functioning, interpersonal relationships, psychotic symptoms, alcohol/drug use, and emotional lability. The measure has demonstrated acceptable test-retest reliability and internal consistency and good construct and discriminant validity (Eisen et al., 2004). Further studies have supported the reliability, concurrent validity, and sensitivity of the BASIS-24 in specific racial groups (Eisen et al. 2006). A copy of the BASIS-24 is provided in the appendix.

Beck Depression Inventory - II (BDI-II). All patients will complete the BDI-II (Beck, Steer, & Brown, 1996) before each session. The BDI-II is a 21-item, self-report questionnaire designed to assess recent depressive symptoms. Each item is rated using a 4-point severity scale, ranging from 0 to 3. The total score ranges from 0 to 63 with greater scores suggesting greater depressive symptoms. The BDI-II has shown good test-retest reliability (r = .93) and has demonstrated a high correlation with the original BDI (r = .93, Beck et al., 1996). A copy of the BDI-II is attached.

Quality of Life Inventory (QOLI; Frisch, 1992). The QOLI is a 32-item self-report measure which first rates the importance of something in a persons life, such as money or self-esteem, on a 3-point scale, and then rates how satisfied a person is with this item on a scale ranging from -3 to 3. A copy of the QOLI is provided in the appendix.

The Medical Outcomes Study 36-item Short-Form (SF-36; Ware & Sherbourne, 1992). The SF-36 is a widely used standardized, generic self-report of health status for evaluating physical and mental health-related quality of life. The SF-36 consists of 36 items; 35 of the items group into eight multi-item scales that collectively measure health-related quality of life: Physical Functioning (10 items), Role Physical (4 items), Bodily Pain (2 items), General Health (5 items), Vitality (4 items), Social Functioning (2 items), Role Emotional (3 items), and Mental Health (5 items). The remaining item concerns the experience of change in general health during the last year. The 36 items are measured on Likert-type scales and the response choices vary between two and six levels. The scores from the eight scales are transformed, with each scale ranging from 0 (worst health) to 100 (best health); the item scores are averaged to create each scale score. After the eight scores are weighted, two summary scales, the Physical Component Summary and the Mental Component Summary, are calculated and standardized to a mean value of 50 and a standard deviation of 10. A copy of the SF-36 is provided in the appendix.
Treatment Expectations Form (Moras & Jones, 1992). This is an adaptation of a form used in the NIMH Treatment of Depression Collaborative Research Program (Elkin et al., 1989). This form is completed at intake and will be used to determine whether treatment groups were matched in terms of pre-treatment expectations for benefit. A copy of the Treatment Expectations Form is provided in the appendix.

Treatment Credibility Form: Opinions About Treatment (Borkovec & Mathews, 1988). This questionnaire is administered early in the treatment to obtain the patient's opinion of the probable value of the treatment he/she is receiving for his/her problems. It will be used to determine whether the treatment approaches are equal in credibility. A copy of this form is provided in the appendix.

The Inventory of Interpersonal Problems (IIP; Barkham et al., 1996; Kim & Pilkonis, 1999). The IIP is a 48-item self-report measure designed to assess a variety of characteristics related to interpersonal relationships, including possible personality disorders. This measure was created by the research team by combining the IIP-25 (Kim & Pilkonis, 1999) with the IIP-32 (Barkham et al., 1996) and consolidating the 9 items shared by the two versions of the IIP.

The Working Alliance Inventory Short Form (WAI-S; Tracey & Kokotovic, 1989). The WAI-S is a 12-item self-report measure that evaluates the therapist-client relationship. The WAI-S has demonstrated a high degree of internal consistency for all subscale scores and the total score as well as high intercorrelations with the original WAI across all subscales and the total score (Busseri & Tyler, 2003). For the purposes of this study, the WAI-S will be administered only to patients in order to ascertain the patients perspective on the therapeutic alliance. Patients will complete the measure after their 2nd, 4th, 6th, and 8th therapy sessions.

Traumatic Life Events Questionnaire (TLEQ; Kubany et al., 2000). The TLEQ is a 24-item self-report measure designed to assess exposure to a broad range of traumatic events. The TLEQ assesses physical, sexual, and emotional trauma across a variety of specific events, and was shown to produce similar levels of disclosure to structured-interview inquiries of trauma (Kubany et al., 2000). This measure will be administered at the baseline assessment.

Alcohol Use Disorders Identification Test (AUDIT; Saunders et al., 1993). The AUDIT is a 10-item self-report questionnaire assessing alcohol consumption, drinking behavior, and alcohol-related problems. For the purposes of this study, the research staff has expanded the original AUDIT to include questions about the use of other substances. Primary question wording was maintained, but the additional items now refer to the use of illegal substances or the use of prescription drugs in a manner different from that prescribed. The AUDIT will be administered at the baseline assessment.

Dysfunctional Attitudes Scale (DAS; Weissman & Beck; 1978). The DAS is a 40-item self-report measure, in which consumers assess the degree to which they endorse underlying depressotypic assumptions on a 7-point Likert scale. This measure will help to assess mechanisms of change during treatment, and will be administered at baseline, Month 1, and Month 2.

Ways of Responding Questionnaire (WOR; Barber & DeRubeis, 1992). The WOR consists of eight different stressful scenarios presented to consumers along with an initial depressotypic automatic thought. Consumers are asked to state their feelings, thoughts, and possible reactions to the given scenario. Each response is then rated by independent judges for the
presence of a list of possible positive and negative cognitive strategies used to respond to the automatic thought. For this study, a community friendly version of the WOR was developed, containing scenarios and language designed to reflect the experience of community mental health consumers. Consumers are also asked to complete a more straightforward self-report measure which is comprised of 65 items detailing various positive and negative cognitive strategies for handling stressful situations. Consumers are asked to report the degree to which they employ each of these strategies on a 4-point Likert scale. For this study, consumers will complete both the community-friendly scenarios and the self-report version of the WOR.

Psychological Distance Scaling Task (PDST; Dozois & Dobson, 2001). The PDST is a measure of cognitive organization that is thought to capture the presence of underlying depressogenic schemas. Consumers are asked to categorize 80 adjectives on two dimensions valence and the degree to which the adjective applies to them. The manner in which individuals organize adjective content on the PDST is assumed to reflect the degree of schema consolidation or interconnectedness of self-relevant information. No copy of the PDST is available for review because it is a computerized assessment that cannot be uploaded to this site.

Self-Understanding of Interpersonal Patterns Scale Revised (SUIP-R; Connolly et al., 1999). The SUIP-R is a 28-item self-report inventory designed to capture consumers levels of self-understanding of their own unique impairing relationship conflicts. Each item on the scale represents an interpersonal pattern that an individual might experience in his/her relationships, which consumers are asked to rate on a 6-point self-understanding scale. The SUIP-R demonstrates good internal consistency (Cronbachs alpha = .92) and discriminant validity, correlating less than .10 at baseline with measures of symptoms, interpersonal distress, and quality of life.

Administration of Study Instruments
Patients will complete measures at baseline, months 1, 2, 4, and 5 into treatment, and weekly. Baseline measures include the QIDS, SCID-I, BASIS-24, BDI-II, HAM-D, QOLI, SF-36, demographic questionnaire, treatment expectations, IIP, TLEQ, AUDIT, DAS, WOR, PDST, and SUIP-R. Month 1, 2 assessments include the BASIS-24, BDI-II, HAM-D, QOLI, SF-36, IIP, DAS, WOR, PDST, and SUIP-R. The Month 4 assessment includes the BASIS-24, BDI-II, HAM-D, QOLI, SF-36, and IIP. Patients will fill out the BASIS-24 and BDI-II weekly. In addition, the credibility measure will be completed at week 2, and the WAI will be completed at week 2, 4, 6, and 8. All measures are self-report with the exception of the HAM-D, and SCID, which will be administered by trained research staff. Furthermore, we will review patient's charts to track medication usage, as this may influence treatment outcome and retention. All measures and datasets will use patient codes – not patient names or other blatant identifiers.

8. Risk/Benefit Assessment

8.1. Risks
The risks to the consumers include the discomfort of revealing personal information on the self-report and interview-based assessments, as well as during therapy sessions, and the potential for treatment ineffectiveness. Within the scope of psychotherapy and assessment, patients may feel discomfort when disclosing personal information to the therapist or assessor. Study patients also risk nonresponse or limited response to treatment, but patients are always free to end participation in the study and pursue alternate treatment if they so choose. Additionally, consumers will spend time completing the measures. Beyond this, risks are minimal, and generally similar to the risk inherent in receiving treatment as usually provided at the NHS sites. The main risks to therapists are the time needed to participate in the protocol (i.e., workshops
and supervision) and the potential of feeling uncomfortable when receiving feedback during supervision.

8.2 Benefits
Patients who receive CT or SE might experience greater improvement in their symptoms of depression. In addition, there is the indirect benefit of contributing to knowledge about treatment of depression. Clinicians participating in the study may benefit from increased training and regular supervision.

9. Subject Confidentiality
Confidentiality will be scrupulously maintained by standard procedures including the use of participant numbers/codes instead of names, the storage of all data including audiotapes in locked cabinets or rooms, and the withholding of all participant information from release without the express written consent of the individual. Data will be entered into Microsoft Access screens which are password protected. All research staff will be trained in HIPAA regulations, and no patient names will be used on research forms that are transmitted to the University. A business agreement between Penn and NHS will bind all participating staff to HIPAA regulations.

Subject Privacy
Confidentiality is strictly maintained by standard procedures including administering the questionnaires in a private office, use of subject numbers/codes instead of names on data forms, storage of data in locked file cabinets and rooms, and withholding release of all participant information without the express written permission of the participant. Research personnel have completed human subjects and HIPAA training and understand the importance of maintaining subject privacy. No personal identifiers (e.g., name, address) are entered into the computerized dataset, and the consumer's name is not written on any CDs or other media storing the audio recordings. The data, containing only randomized patient numbers, will be entered into Access databases on a secure server. Only authorized research personnel have access to the server. In addition, the Access databases will be password protected. A log allowing us to connect the patient number with PHI will be kept in a locked filing cabinet. Audio recordings will be stored on an external hard drive which will be locked up in a locked filing cabinet at the conclusion of the work day. CD copies of sessions and assessments will be made as additional back-up. These will also be locked up at the conclusion of the work day. Research therapy supervisors will complete the CITI human subjects course and will have to sign out CDs of sessions. A log will be kept, and a research assistant will track the lent CDs. Supervisors will agree to keep the CDs in their sole possession while they are signed out.

10. Compensation
Consumers will be compensated after each individual assessment: Baseline ($50), Month 1 ($50), Month 2 ($50), Month 4 ($25), and Month 5 ($50). (The baseline, month 1, month 2, and month 5 assessments provide larger compensation due to the increased subject burden at these assessments.) Compensation will be given at the conclusion of each individual assessment. Therapists will be compensated $300 for attending the training workshop (or equivalent individualized training). All therapists in both treatment conditions will each receive $25 (their normal clinic reimbursement rate) for attending each of the twice monthly group supervision sessions across the training and randomization phases and any individual supervisions during training. Finally, all therapists will receive an honorarium of $150 for every 2 study clients seen for at least one session. These payments will be added onto their paychecks on their regular payroll cycle schedule.

11. Data and Safety Monitoring
In order to ensure the safety of participants and the validity and integrity of the data, a Data Safety Monitoring Board (DSMB) will be convened, whose chief function will be to ensure safe and effective conduct of the study. The DSMB will be composed of senior investigators at the University of Pennsylvania with experience with intervention studies in the community as well as studies of major depressive disorder. The primary goals of the DSMB are as follows: 1) To monitor and advise on scientific and ethical issues related to the study implementation for the protection of human subjects, 2) To review and approve the protocol and subsequently conduct annual reviews to determine whether participant safety has been adequately safeguarded, 3) To review procedures and decisions regarding the adequate protection of specific participants when investigators break protocol because of adverse events or clinical deterioration, 4) To review progress to see that enrollment goals have been met, 5) To monitor and advise on ethical issues related to adverse events, 6) To oversee the confidentiality of data, and quality of data collection, management, and analysis, 7) To recommend, if necessary, discontinuation, modification, or termination of the study based upon emerging data (in the study and literature) and evaluation of risk/benefit ratio, and 8) When possible, to serve as the final arbiters of whether individuals should be removed from the protocol. Although our treating clinicians are empowered to take whatever immediate action is necessary to safeguard the welfare of individuals, the DSMB will be called upon whenever possible to render judgments in the advent of serious suicidal intent or clinical deterioration. The DSMB will meet once per year, or as needed, in Philadelphia. In advent of emergencies, the DSMB will meet via teleconference. For each meeting, the DSMB will first meet in an open session attended by the principal investigator. The group will first review the research protocol and plans for data and safety monitoring. This group will be used to review any problems in implementing the safety plan and for suggesting any necessary modifications to the safety plan. The DSMB will then meet in a closed session for the purpose of reviewing emerging trial data. Confidentiality will be maintained by providing data without any identifying information to the committee. At the conclusion of the meeting, the DSMB will make recommendations to the PI and the University Internal Review Board (IRB). The DSMB will make recommendations concerning the continuation or conclusion of the study. The DSMB will monitor both safety and outcome data as part of the yearly review. Outcome evaluations will include review of data quality and timeliness, participant recruitment, accrual and retention, and review of interim "masked" outcome results on primary and secondary efficacy measures. Safety evaluations will include review of adverse events and weekly symptom measures for each participant. The DSMB will further consider external factors such as scientific and therapeutic developments that may impact the safety or the ethics of the study. All serious adverse events will be reported to the members of the DSMB and the university IRB within 24 hours. A report of all non-serious adverse events will be provided to the board members and the IRB yearly. NIH will be informed of all actions taken by the IRB as part of the continuing review. Monitoring of the DSMB will be the responsibility of the university IRB. The investigators will insure that the Data and Safety Monitoring Plan is reviewed and approved by the IRB before the initiation of the study protocol. The DSMB will then provide a report to the IRB following each yearly meeting, including recommendations concerning continuations or conclusion of the study.

12. Informed Consent Process
For consumers, informed consent for the baseline assessment will occur before this assessment. The subject will read over the form and have the opportunity to ask questions. The Research Assistant will be available to answer these questions. If the patient prefers, he or she will be able to take the consent form home and discuss it with friends/family. Once the subject has signed the consent form, he or she will be given a copy of it to keep for his or her records. If the subject qualifies for the study after the baseline consent, the consenting process will proceed for the main study and HIPAA consents just as it did for the baseline consent. The
Research Assistant will make it clear that participation is voluntary. Over the course of participation in the study, the Research Assistant will check in with the subject (during assessments) to ensure the subject's participation is going well and that the subject wishes to continue. Therapists will also give informed consent. They will review the consent form with research study personnel, have an opportunity to ask questions and receive answers, and they will receive a copy of the form to keep for their records.
Formal Amendments

- We added the Reasons for Ending Treatment Questionnaire, which we planned to administer once a patient has voluntarily withdrawn from therapy in the study. The exact language we included and received approval from IRB is the following: In the event that a consumer decides to terminate treatment prior to completing the 16 sessions designated by the study protocol, the consumer will complete the Reasons for Ending Treatment Questionnaire. Participants will give consent to complete the Reasons for Ending Treatment Questionnaire either at baseline assessment by signing the Training Phase Consent Form or Randomized Phase Consent Form, or, if they have not given consent at baseline assessment, over the phone prior to completing the questionnaire via the consent script. The Reasons for Ending Treatment Questionnaire will be completed at the consumer's convenience either over the phone or in person during a Month Assessment. This amendment was approved by IRB on 9/27/2011.

- We submitted a modification to clarify the language on the protocol regarding which measures patients would be asked to complete at the month 5 assessment. Nothing was changed on the patient consent form. We simply wanted to clarify on the protocol that the month 5 assessment measures were identical to the ones administered at month 1 and month 2 assessments. This amendment was approved by IRB on 11/22/2011.

- A modification was made to change the suicide exclusion criteria from a suicidal gesture within the past 6 months to a suicidal gesture within the past 3 months. The clinic director, intake staff, and corporate director at our clinic site all agreed that the 6-month window was leading to the exclusion of a large number of potential participants who are no longer presenting with suicidality and who would otherwise be appropriate for the study. After much experience evaluating these consumers for eligibility for the study, the research staff and the safety committee at Northwestern Human Services felt confident that changing the exclusion criteria to a suicidal gesture within the past three months would ensure the safety of clients and allow more eligible consumers who are no longer suicidal to participate in the study. This amendment was approved by IRB on 4/23/2012.

- A modification was made to add a therapist questionnaire for therapists to fill out after they finish training in SE or CT therapy. The questionnaire includes questions about some demographic information as well as the therapists’ experiences while being trained in SE or CT therapy for the study. This amendment was approved by IRB on 8/3/2012.