

THE DEVELOPMENT OF A PSYCHOANALYTIC OUTCOME STUDY: CHOICES, CONFLICTS, AND CONSENSUS

This article tells the story of the development of an outcome study of psychoanalysis and describes the debate that took place over critical methodological issues. The protocol committee included career psychotherapy researchers who have conducted rigorous outcome studies, clinical psychoanalysts, study methodologists, and a statistician with clinical trial expertise. The committee worked for two years to develop the study design. This project is based on the premise that clinical psychoanalysis is a treatment. Areas specifically addressed are the goals and hypothesis of the study, inclusion and exclusion criteria, choice of psychotherapies as comparison treatments, definition of treatments and selection of therapists, use of medication, development of a treatment adherence measure, randomization of patient assignment vs. patient self-selection, and primary outcome measures. The execution of this outcome study will require significant effort and resources. A positive result would boost the standing of psychoanalysis, but the results may not support the primary hypothesis that there are therapeutic benefits unique to psychoanalysis and that psychoanalysis can effect demonstrable changes in a patient's mental life and adaptation that are not achieved by treatments of different orientation and/or lesser intensity. However, more important than whatever specific results emerge is what executing such a study requires of our field: the process of addressing the clinical issues that a study design requires, the creation of a network of analysts around the country working on a common project, and the joining of the clinical psychoanalytic community with a community of psychodynamic researchers.

This article tells the story of the development of an outcome study of psychoanalysis and describes the debate that took place over critical methodological issues.

The story began when a representative of the executive committee of the American Psychoanalytic Association approached me (SR) with a request that I submit a proposal for an outcome study of psychoanalysis for the executive committee's consideration. I replied that no one person could develop such a complex protocol. I suggested that with financial support from APsaA, I could bring together the expertise necessary to develop a psychoanalytically sophisticated and methodologically rigorous study. Though the executive committee was interested in this proposal, they declined to offer financial support. However, there was enthusiasm for the project, and Lynn Moritz, then president-elect, reached out to encourage me to find a way to develop the protocol.

I am a career researcher in the Department of Psychiatry at Columbia University and have designed and been the principal investigator of clinical trials of treatments for depression, including late-life depression and depressed patients with cardiovascular disease. Thus, I am familiar with the expertise, process, and financial support needed to develop a protocol to address complex problems in psychotherapy or pharmacological outcome studies. To pursue this project, I secured a competitive grant from a fund within the department that gives awards to innovative projects including protocol development.

THE PROTOCOL COMMITTEE

Having secured the funds, I next had to recruit a protocol development committee. The committee needed to be large enough to provide the necessary expertise, but small enough to function as an effective working group. Three areas of expertise were essential. First, the committee needed career psychotherapy researchers who have conducted rigorous outcome studies.

Steven P. Roose, Professor of Clinical Psychiatry, Columbia University College of Physicians and Surgeons; Research Psychiatrist, New York State Psychiatric Institute; Chair, Research Committee, Columbia Center for Psychoanalytic Training and Research. Madeleine Terry, doctoral student and graduate assistant, Department of Clinical Psychology, City University of New York. The Psychoanalytic Outcome Research Committee: Jacques P. Barber, Eve Caligor, Robert J. DeRubeis, Andrew J. Gerber, Robert A. Glick, Mark J. Hilsenroth, Andrew C. Leon, Kenneth N. Levy, Raymond A. Levy, Barbara Milrod, Bret R. Rutherford, Katherine Shear, Michael E. Thase, Drew Westen.

Supported by a grant from the Dorgan Fund, Department of Psychiatry, Columbia University College of Physicians and Surgeons. Submitted for publication October 17, 2011.

As evidence of this level of expertise, they should have received NIMH support for their studies. These researchers would provide the methodological knowledge and experience required to address such problems as which treatments to compare with psychoanalysis, what population to target, assessment of personality dimensions and defense mechanisms, the training of therapists, assessment of treatment adherence, and the critical issue of outcome measures. Second, the committee required the input of clinical psychoanalysts. Third, the committee needed a statistician with clinical trial expertise. In outcome studies, statisticians are not simply brought in at the back end to analyze the data; they need to be included from the beginning to help construct the study design and ensure that the study hypothesis can be tested with the data collected from the sample size projected.

In forming the protocol committee, it was also important to include young researchers at the beginning of their careers; an outcome study of psychoanalysis will require at least a decade to complete, and it is the responsibility of senior researchers to engage and train the next generation of researchers to carry the work to completion. In addition, it was important to engage APsaA through the participation of Dr. Moritz, by now president, in developing the protocol; this was necessary to convince the leadership of organized psychoanalysis that the study would be important to the field.

I consulted psychoanalytic and research colleagues for suggestions regarding who might serve on the committee. All of the people I contacted were fascinated by the challenge and were eager to join the group. The list of those who participated in the protocol development appears in Appendix A.

BEGINNING THE WORK

Once the committee was formed, participants were assigned to subcommittees to prepare for the first working session. The areas covered by the subcommittees were (1) goals and hypothesis; (2) inclusion and exclusion criteria; (3) the comparator treatment to psychoanalysis (e.g., a psychotherapy treatment, medication, or a treatment as usual group) and whether the study should be a randomized controlled trial (RCT) or a naturalistic comparison; (4) definitions of treatments (e.g., will we require a treatment manual?); (5) treatment adherence; (6) outcome measures, projected effect sizes, and power analyses to determine the number

of patients needed. Each subcommittee was instructed to formulate the critical questions, propose methodology for the study, and specify readings that would inform the rest of the committee. The subcommittees were required to circulate their proposal and reading list at least two weeks before the first meeting.

Developing a study design for an outcome study of psychoanalysis has significant educational importance for the field, even if the study were never to be executed. As one can infer from the subcommittee topics, when designing a clinical trial one is forced to address several crucial questions in everyday clinical work: To whom do we recommend psychoanalysis? What defines psychoanalytic treatment? What are the criteria for termination? How do we know if there has been therapeutic benefit?

The committee held its first meeting over the course of a weekend in the spring of 2007 in New York. The committee met all day Saturday and until one o'clock the next afternoon. The work continued by e-mail and conference calls over the next eight months, and at a second weekend meeting in the fall of 2008 the protocol was completed.

The protocol was then submitted to Dr. Moritz, who asked Linda Mayes, then chair of the Fund for Psychoanalytic Research, to select and chair a special review committee. Mayes selected three reviewers, who remained anonymous. The reviewers required two resubmissions before approving the protocol and recommending funding. The protocol committee felt that the review process was fair, rigorous, and constructive.

THE PROTOCOL

Goals and Hypothesis

This project is based on the premise that, as stated on the home page of the website of the American Psychoanalytic Association, clinical psychoanalysis is a treatment. The first task was to articulate the goal of the study. The goal and the hypothesis of a study are related but nonetheless different. The goal is what we hope to accomplish by doing the research and in this case reflects the belief of psychoanalysts that psychoanalysis changes people in profound, meaningful, and enduring ways that only a treatment of this depth and intensity can accomplish. Thus, the goal of the study is to demonstrate that there are therapeutic benefits unique to psychoanalysis, and that psychoanalysis can effect demonstrable changes in a patient's mental life and adaptation that are not achieved by treatments

of different orientation and/or of lesser intensity. Though psychoanalysts may believe that psychoanalysis will uniquely benefit some patients, in reality there are many other clinical treatments available. It is therefore insufficient to simply demonstrate that psychoanalysis is as effective as other treatments with respect to symptom improvement; rather, it is necessary to demonstrate that there are robust, sustained, and meaningful, indeed life-changing, therapeutic benefits that occur only in psychoanalysis. This unique therapeutic gain justifies the time and expense necessary to support the intensity and duration of psychoanalysis.

The hypothesis of a study translates the goal of the study into a statement that will or will not be supported by the results. The primary hypothesis of this study is that patients treated by psychoanalysis will have significantly greater improvement in interpersonal relationships, decreased personality pathology, and greater improvement in work and play compared to patients receiving treatments of a different theoretical orientation or lesser intensity. The study was designed to determine whether there is added value in psychoanalytic treatment compared to less intense psychotherapies. The goal of the study clearly represented the beliefs of the psychoanalysts on the committee. Not surprisingly, other members—dynamic and cognitive-behavioral psychotherapy researchers—did not share these beliefs. However, given that the hypothesis was cogent and the study design unbiased, the goal of the study was not a concern for these nonpsychoanalytic members of the committee. A methodologically rigorous study design will produce results that support or do not support the hypothesis. Once there was a testable hypothesis, the committee turned to the challenging issues of study design.

Inclusion and Exclusion Criteria: What Type of Patients Should Be Included?

Inclusion and exclusion criteria define the patient sample to be included in the study. Constructing these criteria requires an answer to a long-standing but unanswered clinical question: Who are the patients that psychoanalysts recommend to receive psychoanalysis in preference to other treatments? Why analysts recommend psychoanalytic treatment for some patients and not others is a long-discussed but little-researched topic. When an analyst recommends psychoanalysis as the treatment of choice to a patient, we infer that this recommendation is based on the expectation of a positive outcome—that this type of person with this type of problem will benefit, perhaps uniquely so, from psychoanalysis.

However, because there are no controlled outcome studies of psychoanalysis we lack an empirical basis on which to make the recommendation that psychoanalysis is indicated for one patient but not for another.

In the absence of systematically collected data, it is customary and reasonable to rely on clinical consensus. However, as with many important concepts in psychoanalysis (e.g., the definition of psychoanalytic process), there is no clinical consensus about which patients should be recommended for analysis. Some believe the decision should be based on diagnosis (e.g., that patients with higher-level personality disorders should be treated with analysis, whereas those with severe personality disorders or substance abuse should not be [Kernberg 1999]). Others believe the recommendation for psychoanalysis depends not on diagnosis but on the presence of particular psychological functions, for example, ego strength or psychological mindedness (Wallerstein 1994). Yet another group suggest that it is not possible to predict who will engage with and benefit from psychoanalysis, and therefore recommend a “trial of analysis” for all who are willing (Rothstein 1990).

With neither clinical consensus nor outcome studies to guide them, protocol committee members decided to construct inclusion and exclusion criteria based on available data from recent studies that systematically assessed patients recommended to enter psychoanalysis after a standard clinical evaluation. The largest and most methodologically rigorous study describing patients recommended for analysis (Caligor et al. 2009) included patients applying for psychoanalytic treatment as training cases at the Columbia Center for Psychoanalytic Training and Research. Patients were assessed using standard psychometric measures and then received a standard clinical assessment of three to eight sessions done by a supervised psychoanalytic candidate. The decision to recommend psychoanalysis or not was based exclusively on the clinical evaluation and not on the research data, which was not made available to the candidate or supervisor doing the evaluation. There was a high level of psychiatric morbidity and psychosocial impairment in these patients. With respect to DSM-IV Axis I diagnoses, 50% had a current and 74% a lifetime diagnosis of mood disorder, and 56% a current and 61% a lifetime history of anxiety disorder. With respect to current symptoms, the mean Beck Depression Inventory score (a self-report measure used to assess symptoms in patients with mood disorder) was in the moderately severe range, 19.1 ($SD = 11.0$); the mean Hamilton Depression Rating Scale score (traditionally used in antidepressant clinical trials) was in the mild range,

14.1 ($SD = 7.8$); and the mean Hamilton Anxiety score was in the moderate range, 14.6 ($SD = 8.1$). Fifty-seven percent met criteria for an Axis II diagnosis, and the mean Social Adjustment Score (an evaluation of problems in work, leisure, and family and personal relationships) indicated moderate to high impairment consistent with the rate of Axis I and II disorders.

Thus, in the Caligor et al. study, the results of which were replicated in a second sample, patients recommended for psychoanalysis had chronic symptoms of depression and anxiety, high rates of mood and anxiety disorders, and significant dysfunction in interpersonal relationships. These data were operationalized into the following inclusion criteria for the proposed study:

- (1) Men and women, age 18 to 55
- (2) DSM-IV depressive disorder: either major depressive disorder, dysthymia, or depression NOS
- (3) Baseline Beck Depression Inventory score between 14 and 26
- (4) Depression diagnosis present for at least one year, or six months if there has been a prior episode of major depression or dysthymia
- (5) A minimum score of 1 on the Inventory of Interpersonal Problems

The intent was to include patients with relatively stable symptoms so that subsequent change could be attributed with greater confidence to treatment response.

The exclusion criteria were somewhat easier to formulate, as there is some consensus among psychoanalysts that patients who are acutely psychotic or suicidal, have chronic and significant self-injurious behavior, or have current substance dependence (excluding nicotine) are generally not recommended for psychoanalysis. The exclusion criteria are:

- (1) Acutely psychotic or history of psychotic disorder
- (2) Imminent need for hospitalization
- (3) Chronic, physical self-destructive behavior
- (4) Substance dependence other than nicotine
- (5) Acute, severe, or unstable medical condition
- (6) Taking psychotropic medication if the dose has not been stable for three months

It was debated whether a determination should be made regarding the patient's being "analyzable" or "suitable for analysis." This proposal was rejected because there is no consensus among analysts on how to operationalize these terms, nor are there data to support their predictive validity.

What Should the Comparator Treatment or Comparison Group Be for Psychoanalytic Treatment?

The next issue considered by the protocol committee was whether the study should compare psychoanalysis to other active treatments or use a naturalistic study design and compare psychoanalysis to treatment (or no treatment) as usual. It is believed by some that a naturalistic study design has more ecological validity, that is, that the results will reflect the difference between people who have an analysis and those who do not as it exists in clinical reality. The problem is that without control for patient assignment or assessment of what treatment actually takes place, it is hard to interpret the results. By contrast, a well-designed controlled comparator study addresses the critical clinical question: Are there other treatments, of a different theoretical orientation or significantly lesser intensity, that have therapeutic benefit comparable to that of psychoanalysis? Given that the goal of the study is to demonstrate that psychoanalysis has clinically significant, unique therapeutic effects that are not achieved by less intense treatments of shorter duration and/or not psychodynamically based, the committee decided that a comparator controlled study design was necessary.

An obvious comparison treatment was cognitive-behavioral therapy (CBT): it is the most researched psychotherapy treatment, has the largest evidence base to support its effectiveness, and is widely practiced. Moreover, it has the advantage of being different from psychoanalysis not only in duration and intensity but also in its theoretical underpinnings. For many reasons, it is desirable to compare only two treatments in a study (most notably, it makes the study more feasible by limiting the number of patients needed to participate). However, if a significant difference were to be found in a two-treatment comparison between psychoanalysis and CBT, the question would remain whether this difference resulted from the greater intensity and duration of psychoanalysis (four times a week for a minimum of four years, as against CBT once or twice a week for a duration of months, not years), or whether it resulted from differences in theory and technique between the two treatments. The inclusion of a dynamic psychotherapy (DP) treatment cell would deconstruct the issue of intensity and duration versus theoretical approach. Furthermore, a number of recent studies have demonstrated the effectiveness of time-limited DP for treatment of depression and panic disorder (see Milrod et al. 2007; Gerber et al. 2011). Though a three-cell study would increase the size of a two-treatment comparison by 50%, thereby

making the study more difficult to execute, the committee agreed that it was compelling to include a DP cell whose treatment would be identical in intensity and duration to the CBT treatment and to compare both with psychoanalysis.

In any comparative study, all groups should be selected using the same inclusion and exclusion criteria; thus, inclusion and exclusion criteria for patients in the psychotherapy cells must be the same as for those receiving psychoanalysis. The criteria constructed with psychoanalytic patients in mind presented no problem for the committee members representing the psychotherapy comparator groups; they felt that these criteria described a patient population equally suitable for treatment with CBT or dynamic psychotherapy.

Once the psychotherapy comparators to psychoanalysis were decided, the next task was to establish the parameters of the psychotherapy treatments, specifically the frequency of sessions and duration of treatment. The committee debated whether the psychotherapy treatments should replicate the most common conditions in randomized controlled studies of CBT or DP, in which duration is no longer than six months, or whether the psychotherapy treatments should have greater flexibility in order to reflect more real-world clinical conditions. This decision was left up to the psychotherapy researchers on the committee. Should the study find that psychoanalysis has a better outcome than one or both of the psychotherapies, it was hoped that the study design would preclude the protest that the comparisons were biased because the psychotherapy conditions offered less than optimal treatment. Thus, the psychotherapy researchers recommended that both the CBT and the DP treatments be flexibly dosed, one or two sessions a week, with duration up to one year. The total number of sessions would not exceed 46 during the maximum one-year period of treatment.

HOW SHOULD PATIENTS BE ASSIGNED TO A TREATMENT CONDITION?

The most debated and critical decision made by the protocol committee was that patient assignment would be randomized rather than decided by patient choice. There is increasing recognition of the limitations of a randomized controlled trial design (Ware and Hamel 2011; Barber 2009), but as all study designs have strengths and limitations, the task of the committee was to determine the design best suited to test the hypothesis

of the study. Randomization is the accepted method to mitigate the potential impact of our limited knowledge of the many complex variables that may cause differential outcome between treatments if these variables are not evenly distributed among the three treatment conditions. For example, we assume in this study that there may be personality dimensions or ego variables that effect outcome, but we have no evidence of what they might be. Thus, the study relies on the expectation that randomization will evenly distribute the unknown variables across the three treatment groups so they will not differentially impact outcome. In relatively large studies, randomization generally works well, though not infallibly, in the even distribution of known variables and presumably equally well in the even distribution of unknown variables. If we know there is a variable that significantly predicts outcome, then a procedure called stratification is executed prior to the randomization. For example, if it had already been established that women respond better to treatment than men, then gender would be stratified across the three treatment cells so that an equal number of men and women in each treatment condition would be ensured. Currently there is no variable established that predicts outcome, so there is no need to stratify the sample prior to randomization.

An alternative study design to randomization is patient self-selection, that is, allowing patients to choose which of the three treatments they will enter. This design yields important data about the “real-world” acceptability of the different treatments. It also may favor a higher response rate for all three treatments, because patients are getting the treatment they prefer, and this should increase their expectations of improvement. However, the results from a study with this design would provide limited information about the differential effectiveness of the treatments and would not provide information regarding which treatment a clinician should recommend to a patient.

Thus, each study design has its advantages and disadvantages and ultimately answers different questions. The committee did not reach unanimity on this issue. The decision to choose a randomized controlled trial design was based on a majority conclusion that this design was best suited to the primary goal of the study, that is, to compare psychoanalysis to briefer psychotherapy treatments to test whether there is a unique benefit to psychoanalysis. There is another important reason the committee chose an RCT design: the recognition that if a study demonstrates a unique effect of psychoanalysis, the results must be convincing not only to psychoanalysts but also to a wider professional audience. Even with its

acknowledged limitations, the gold standard for comparing treatments is still the randomized controlled trial. Thus, the study design that emerged from this deliberative process was a randomized comparison between clinically optimized, flexibly dosed CBT, DP, and psychoanalysis for the treatment of patients with mild to moderate chronic depression and significant interpersonal problems.

SHOULD PATIENTS ON MEDICATION BE INCLUDED?

Another issue with significant implications that the protocol committee addressed was concomitant use of antidepressant medication. Given the anticipated rate of Axis I mood and anxiety disorders in the target population and the symptom severity of patients entering psychoanalysis, it is not surprising that studies have shown that 30 to 40% of patients in analysis are also being treated with antidepressant medication (Caligor et al. 2009). In considering whether to allow patients in the study to be treated with antidepressants, the protocol committee faced two conflicting needs. It would be preferable to compare the psychotherapies and psychoanalysis without the complicating impact of medication, so that if there were a difference in outcomes the presumptive conclusion would be that the difference was attributable to the differential effects of the three treatments. Allowing a second concurrent treatment (medication) would significantly complicate the comparison of the therapeutic effectiveness of the interventions being studied. Moreover, there is the risk that the rate and/or adequacy of medication use would be unevenly distributed across the treatment cells, further confounding interpretation of the outcome data. However, the desire to keep the psychotherapeutic interventions pure must be weighed against the need to make the study feasible and as close to real-world clinical conditions as possible. To exclude patients who are on antidepressants at the time of evaluation or for whom the treating clinician wants to recommend medication during the course of the study would mean excluding as many as 40% of all patients who met the inclusion criteria. This would markedly reduce the study's feasibility (recruitment of the patient sample) and significantly skew the patient sample, thereby reducing the generalizability of its findings.

To balance these conflicting needs, the protocol committee, agreeing it was necessary to allow medication, constructed guidelines to standardize its use. Patients would be allowed to enter the study if they were on psychotropic medication and the dose had been stable for three months.

Patients would not be started on psychotropic medication for the first three months of the treatment. However, there would be a rescue provision: if in the judgment of the clinician a patient's depression had deteriorated, a referral for a medication consultation could be made before three months. Any patient a therapist feels should be evaluated would be seen by the medication consultant and, if needed, prescribed medication. A single pharmacological psychiatrist for the study would manage the medication for all study patients, regardless of the training of the therapist. Medication treatment would be recommended to a patient if there had been a persistent worsening of depressive symptoms that (1) had been consistent for at least four weeks and (2) represented a 25% increase in the baseline BDI score. It was believed that these guidelines (and the medication consultant can make a recommendation if clinically indicated regardless of the guidelines), along with a medication algorithm, would decrease the variability in medication management. However, insofar as the therapist would make the referral, an unavoidable variability would be possible in the rate of referral across the three treatment groups. The only way to minimize this potential variability would be to set strict criteria for when the therapist could make a referral. This would likely decrease the variability in referral rate but would also compromise good clinical practice. Further, if the therapist believed a patient needed a consultation despite not meeting the referral criteria, the therapist might drop the patient from the study so the patient could receive optimal treatment. For those reasons the committee chose not to strictly control referral for medication and to risk a possible problem in differential referral rate across the three cells. Patients treated with medication would continue in the protocol, and medication would be noted for purposes of outcome analyses. To control for the effect of medication on the impact of the psychotherapies and psychoanalysis, medication status will be covariate in analyses of outcome.

With respect to the algorithm for medication treatment, the best option seemed to be to follow the algorithm used in the STAR*D study (Trivedi et al. 2006). In the proposed study, patients would begin with an SSRI, citalopram. If a patient did not respond to citalopram or had already failed an SSRI, treatment options would include (1) a switch to a second SSRI, (2) a switch to a "dual action agent" such as duloxetine or venlafaxine, or (3) augmentation with bupropion. If the patient still did not respond after a switch and/or augmentation, the results from level 3 and level 4 of STAR*D are not strong enough to warrant including these

steps as part of the algorithm. The medication consultant and the patient would decide further treatment.

SELECTION OF THERAPISTS

Once the psychotherapy comparison groups and the frequency and duration of treatment were finalized, the next issue for the committee to consider was what qualifications to require of the therapists. Committee members representative of each treatment condition proposed how therapists should be selected for their arm, at which point the entire committee discussed the proposal. The psychoanalysts proposed that therapists in their arm must be members of APsaA and have either an M.D. or a Ph.D. Then three possibilities were discussed. The first was that advanced psychoanalytic candidates be included in the study. This would certainly increase feasibility, and the cases would be supervised, but it was decided that relying on beginning practitioners, supervised or not, would not be fairly representative of psychoanalysis. The second possibility was the polar opposite, to include only training analysts. Treatment given by experienced analysts recognized for their clinical excellence would indeed ensure that psychoanalytic treatment was optimally represented, but it would be a skewed representation. Most practicing psychoanalysts are not training analysts, and this restriction would not only reduce feasibility but would also markedly limit the generalizability of the study's findings. The third possibility was to accept any analyst who is an APsaA member, and this was chosen as best representing how psychoanalysis is currently practiced.

The therapists in the CBT or DP cells are clinicians, M.D.s or Ph.D.s who identify themselves as CBT or DP therapists but do not have analytic training. Therapists for these cells will have had ten years of clinical experience or will already have served as therapists in studies of CBT or DP that used treatment manuals, assessed treatment adherence, and required therapists to meet an adherence standard. Given the intensity and duration of psychoanalysis, there are necessarily going to be more therapists in the psychoanalytic cell than in the psychotherapy cells. In the course of a five-year study, a single CBT or DP therapist may treat fifteen to twenty patients, in contrast to the one or two patients a psychoanalyst might treat. In the psychotherapy cells, the number of patients per therapist could be large enough that therapist differences could be analyzed with respect to outcome; obviously that would not be true in the

psychoanalytic cell. The inclusion of a greater number of therapists in the psychoanalytic condition is a practical necessity. Given the reality that despite any manual or guidelines with respect to technique there will be variability in the approach, style, and technique of the psychoanalysts, the benefit of including more therapists in this group is that if there is a positive finding for psychoanalytic treatment as hypothesized, we can be confident that the result reflects what happens in psychoanalytic treatment generally, as opposed to reflecting the skills of individual clinicians.

DEFINITION OF TREATMENTS

Having decided on the treatment conditions, selection of therapists, and the use of concurrent treatments (medication), the committee then addressed the issues of definition of treatments, use of treatment manuals, and measurement of treatment adherence. With respect to definition of the psychotherapy conditions, the accepted standard for psychotherapy studies is the use of treatment manuals, and there are established treatment manuals for both CBT and DP. The use of manuals optimizes treatment delivery and improves adherence ratings.

There is no treatment manual for psychoanalysis. The definition of a psychoanalytic treatment, including such variables as frequency and duration, as well as what constitutes standard psychoanalytic technique, is obviously complex and has vexed the field since its inception. The committee turned to its psychoanalytic members and, in particular, to the clinical psychoanalysts to address these issues. There are no empirical data available on the impact of session frequency (e.g., three vs. four times a week) or duration of treatment (e.g., a minimum of three years vs. a minimum of five) on outcome. In the absence of data from clinical trials, the next step was to see if a clinical consensus could be found that might serve as a guide. The guidelines promulgated by APsA are taken by many to represent the standards accepted by clinical psychoanalysts. Currently those guidelines set the minimum preferred frequency at four sessions a week, but the committee acknowledged that many analysts see patients three times a week. This was considered an acceptable, but not preferred, option. The patient should lie on the couch. Because psychoanalysts use different techniques in beginning a treatment, an induction phase of up to three months would be allowed before the patient must be seen on the couch three or four times a week. It was recognized

that session frequency might fluctuate between three and four times a week during the course of the analysis.

The clinical psychoanalysts on the committee felt it important to agree on a description of the fundamental tenets of psychoanalytic technique and an accompanying glossary of terms. This description would not be tied to a specific theory of mind (e.g., ego psychology, object relations theory, or self psychology).

Some empirical data do exist on the duration of analyses, if not on its impact on outcome. From studies done at the Columbia Center for Psychoanalytic Training and Research, we know that the mean duration of training cases is 6.1 years (Glick et al. 1996) and of training analyses for Columbia candidates is 6.6 years (Cherry, Wininger, and Roose 2009). Data from other sources, including APsaA surveys, show a wide range of treatment duration. Since it seemed unreasonable to arbitrarily define a minimum duration, it was decided that at the end of each analysis, the analyst would report whether the analysis ended prematurely or had reached a therapeutic termination.

TREATMENT ADHERENCE

The issue of treatment definitions and manuals is related to the issue of adherence. All studies, whether researching psychopharmacology or psychotherapy treatments, need to demonstrate that the treatment being studied is in fact the treatment being delivered. In pharmacology studies, the issue of compliance is equivalent to adherence, and whether patients are taking their medications is assessed by plasma levels and pill counts. In psychotherapy studies adherence is addressed by direct assessment of the sessions. The standard methodology is that all sessions are audio- or videotaped so that patient and therapist become acclimated to the taping; a series of sessions are then randomly selected for the analysis of adherence rating. An adherence measure includes an assessment of core elements that should be present in the treatment, as well as elements that should not. There is already an established adherence measure that is used to differentiate CBT from DP (Hilsenroth et al. 2005). A member of the committee, Mark Hilsenroth, is one of the creators of this measure and adapted it to include adherence questions from other reliable adherence measures of CBT (Hollon et al. 1988; Strunk et al. 2007), as well as ideal technique ratings from senior training analysts (Ablon and Jones 2005). Further, we directly consulted with

clinicians and supervisors from all three cells to ensure that items on the measure are relevant to the actual practice of each of the treatments. In sum, our method in developing the adherence scale for the COPPS project included (1) choosing relevant items from existing measures with demonstrated reliability; (2) reviewing the research process literature to develop related items; (3) consulting with clinicians and supervisors in each of the cells. Finally, we plan to conduct a psychometric evaluation of this scale based on data from pilot study sessions and make empirically based additions, deletions, or modifications.

Although measurement of adherence is standard in psychotherapy studies, the analysts on the protocol committee raised the issue of whether assessment of psychoanalytic treatments should focus on the development of an analytic process rather than on adherence to psychoanalytic technique. Psychoanalytic process is a compelling concept, but there is a long history of disagreement among analysts about its definition, let alone how to assess its presence (Vaughan et al. 1997). It was therefore decided that rating psychoanalytic treatments on the basis of analytic process was not feasible.

The importance placed on measuring adherence stems from the need to ensure that the treatment provided conforms to the intention of the treatment. It stems also from the expectation that adherence is positively and robustly associated with therapeutic outcome. However, with respect to psychoanalysis there is debate about the therapeutic importance of adherence. Some psychoanalytic clinicians maintain that pivotal moments in an analysis often happen in the context of a deviation from standard technique, a shift, however temporary, in the therapeutic frame. They emphasize that flexibility, spontaneity, and genuineness of response in the moment may be as important as adherence to standard psychoanalytic technique (Barber 2009). The DP and CBT clinicians in the study emphasized that these factors and adherence to a treatment model are not mutually exclusive, nor does adherence foreclose an optimal responsiveness to the patient and the ongoing process of the session. Measuring adherence in this study makes it possible to address this clinical question empirically. It may well be that in psychoanalysis the highest adherence scores inversely correlate with therapeutic gain, or some other variation of these two variables could be the case. Again, the point of measuring adherence is not to enforce orthodoxy, but rather to demonstrate that the three treatment conditions in the study are different in more than name only, as well as to examine the relation between technique and outcome.

OUTCOME MEASURES

The proposed study includes patients with chronic depression and difficulties in interpersonal relationships and randomizes them to one of three treatments, CBT or DP lasting up to a year or psychoanalysis three or four times weekly on the couch for an unspecified duration. The primary hypothesis of the study is that patients in psychoanalytic treatment will have greater and more sustained improvement in interpersonal relationships and function at the end of treatment compared to the other treatment groups at the end of a year. How will we measure outcomes? A study such as this can measure many variables, all of them of potential interest. However, for the study to have the statistical power to test the primary hypothesis, it was necessary a priori to designate a single or at most two primary outcome measures. These outcome measures will be the primary statistical comparators, and all the other variables that are assessed will be tested in secondary analyses. Failure to take this approach means that a statistical correction for multiple comparisons would need to be applied to the results, causing a loss of statistical power to detect significant differences among the three treatments.

How did the committee choose the primary outcome measures? Since the primary hypothesis of the study is that there is a unique benefit to psychoanalysis, it makes sense that the primary outcome measures assess the domains that psychoanalysts believe will show the therapeutic gains unique to psychoanalysis.

Once a domain is chosen as best representing the unique change possible with psychoanalytic treatment, a second critical issue is whether there is an instrument with demonstrated psychometric properties that measures this domain. After much discussion, the committee agreed that the Shedler-Westen Assessment Procedure (SWAP-II-200) should be one of the primary outcome measures. The SWAP comprises 200 personality-descriptive statements that are arranged by a clinical observer in a fixed distribution to describe a patient (Westen and Shedler 2007; Shedler and Westen 2004a,b, 2007). Using a dimensional model of personality derived from factor analysis, the SWAP enables experienced observers to quantify clinical observations in a consistent, detailed, and comprehensive manner. The SWAP generates a set of scores for sixteen domains including psychological health, obsessionality, schizotypy, emotional avoidance, emotional dysregulation, narcissism, and so on. All the domains include items that address internal psychological processes,

such as ways of experiencing the self and experiencing others. For this study the psychological health domain will be the primary outcome measure. The psychological health trait factor assesses the presence of a patient's positive psychological strengths, capacity for insight, reflection, and empathy, healthy expression of emotion and love, pursuit of goals and accomplishments, and ability to find meaning and contentment in life experiences and personal relationships. In this study, an interviewer will conduct the semistructured Clinical Diagnostic Interview (about two and a half hours long) at baseline and at termination of treatment. A clinician-rater will then review the interviews and generate the SWAP scores.

Because the SWAP has not been widely used in outcome studies, the committee thought it important to choose a second primary outcome measure, one that has been widely used, so that the results of this study can be compared to a broader literature. It was decided that the second measure should focus on psychosocial function, and The Range of Impaired Functioning Tool (LIFE-RIFT) was selected (Leon et al. 1999, 2000). This is a rater-administered measure, and ratings are based on concrete behavioral indications of functioning. This scale, with established interrater reliability, is administered using a semistructured interview that requires some clinical judgment by the interviewer. The LIFE-RIFT assesses psychosocial function across four domains: work, interpersonal relations, recreation, and global satisfaction. A score is generated for each domain, as is a total score.

With respect to the choice of primary outcome measures, the protocol committee strongly differed with other recent outcome studies comparing psychoanalysis to psychotherapies in depressed patients. In these studies (some ongoing) conducted in Europe, a depression severity scale is the primary outcome measure (Huber and Klug 2005). The protocol committee noted that there are many effective treatments for depression, including specific psychotherapies such as CBT, interpersonal therapy, and brief dynamic therapy, as well as antidepressant medications. Since the purpose of the study is not to show simply that psychoanalysis is yet another effective treatment for depression, but that it has a unique, more global therapeutic benefit, a depression scale should not be a primary outcome measure (though in this study it is designated as a secondary outcome measure).

A major methodological issue is how to compare outcomes of treatments of different durations and intensities. This issue is still being debated among the committee, and undoubtedly there is more than one

reasonable approach. One point of view is that you compare treatments as they are clinically given. The researchers who constructed the psychotherapy cells believed that 46 sessions in one year is an optimal representation of their treatments and that the outcome will compare favorably with psychoanalysis requiring four to six years and hundreds of sessions. It is one of the proposed advantages of the psychotherapies that they do not require the same intensity and duration as psychoanalysis. Thus, the clinically correct comparison is the end of psychotherapy treatment with the end of psychoanalysis (as can be seen in the design, we do a first comparison at the end of four years of analysis). However, some committee members think it important to compare treatments at the same point of their duration, though of course matching for duration does not match for intensity. There is a planned evaluation of the analytic patients at one year, but its primary purpose is to track the trajectory of change in this treatment (all the change may occur *after* one year). Thus, a comparison of the three treatments at the end of a year is possible. We also plan secondary analyses of number of sessions attended as a covariate to test dosage effects. The plan is to conduct annual follow-up assessments for four years of all psychotherapy patients who complete treatment. This makes it possible to see if gains in treatment are sustained.

HOW MANY PATIENTS ARE REQUIRED TO TEST THE HYPOTHESIS?

The size of the study—the number of patients needed in treatment to test the primary hypothesis—was determined by a power analysis. The power analysis determines how many subjects must be enrolled in the study so that if a statistically and clinically meaningful difference in the primary outcome measure exists between the psychotherapies and psychoanalysis, the study has an 80% chance of being able to demonstrate that difference. Given (1) data on the magnitude of positive change in psychosocial function after a year of psychoanalysis (Vaughan et al. 1994) and (2) definitions of what would constitute a clinically meaningful difference between the psychotherapies and psychoanalysis, the power analysis determined that 360 patients, 120 in each treatment condition, would be adequate to test the primary hypothesis.

Details of the data analysis plan are beyond the scope of this paper. In summary, the primary data analysis is an intent-to-treat analysis, which means that all patients who are randomized are counted in data

analysis regardless of how long their treatment lasts, (e.g., if after randomization a patient never shows up for treatment or if a patient drops out after three sessions, he or she is included in the data analysis). In addition to the intent-to-treat analysis, there will be an outcome analysis of patients categorized by their analyst as reaching a therapeutic termination compared with patients who complete their psychotherapy treatments. This is the so-called “completer” analysis. Because it will be too long to wait until all the analytic cases have reached termination, a data analysis is planned once all analytic cases have reached the four-year mark. At that point data from the three treatment conditions will be compared.

THE PILOT STUDY

When the study design and the specifics of the protocol were finalized, the next question was whether this study could be executed as proposed. Before attempting to raise money through grant support from foundations and/or from the NIMH to conduct a clinical trial, it is necessary to demonstrate the feasibility of the study design by collecting pilot data. The most important question of feasibility in this study is whether patients will be willing to accept randomization when one of the treatment conditions is psychoanalysis, a treatment that has much greater session frequency and much longer duration than the comparison treatments. In 2009 APsaA’s executive committee pledged \$150,000, augmented by \$50,000 from the Columbia Center for Psychoanalytic Training and Research, to fund a pilot study to demonstrate the feasibility of a randomized outcome study of psychoanalysis compared with CBT and DP. The pilot study has been completed, and the results appear in Caligor et al. (2012).

THE NEXT STEP

With the completion of the pilot study demonstrating the feasibility of the study design, the next step is to raise the funds to execute the full 360-patient study. The study will require participation by six centers nationwide that can deliver all three study treatments. This would make patient recruitment feasible and would make possible the analysis of site variability in the results. The major cost in the proposed study is the cost

of treatment. If cost per session is set at \$100, treatment costs for the entire study (based on the assumption that patients complete their treatment) come to \$1,104,000, or \$221,000 annually, for the two psychotherapies together and \$11,040,000, or \$2,208,000 annually, for psychoanalysis (based on 4 sessions a week, 46 weeks a year, for 5 years). The projected budget for five years is \$16,000,000, with more than two-thirds of that going to pay for psychoanalytic treatment. If there were no cost for the psychoanalytic treatment, the projected budget would be \$4,730,000, or \$946,000 a year.

An annual budget of over three million dollars, two-thirds of which supports the psychoanalytic treatment, is untenable. To do this study, what is needed is for 120 psychoanalysts, evenly distributed across the six treatment centers, to donate their time to treat one case in the study. This would make the budget comparable to other funded treatment studies with respect to cost per patient. The support of APsA and the commitment of 120 analysts to contribute their time will significantly increase, though not guarantee, our chances to secure funding from foundations or the NIMH. Introducing a variable such as differential payment of therapists is undesirable, but it is a necessary evil if there is to be a realistic chance to secure the necessary resources. It should be noted that there is no empirical evidence demonstrating that differential payment of therapists affects outcome, but most methodologists would nonetheless prefer not to introduce such a difference between the groups.

Clinical psychoanalysts have often been skeptical of the value of outcome research, and to help with the next steps, an advisory group of nationally known and respected psychoanalysts was formed (see Appendix B). This group has already helped adapt the study design to the clinical situation. Their advice will be indispensable in the campaign to recruit the volunteer analysts needed to make the study possible.

Execution of this outcome study will require significant effort and resources. Is it worth it? A positive result would boost the standing of psychoanalysis, but the results may not support the primary hypothesis. However, what is more important than whatever specific results emerge is what executing such a study requires from our field: the process of addressing the clinical issues that a study design requires, the creation of a network of analysts around the country working on a common project, and the joining of the clinical psychoanalytic community with a community of psychodynamic researchers. These actions will transcend the final data.

**APPENDIX A: THE PSYCHOANALYTIC OUTCOME
RESEARCH COMMITTEE**

JACQUES P. BARBER, PH.D., ABPP, Professor of Psychology in Psychiatry, University of Pennsylvania School of Medicine and Philadelphia VA Medical Center.

EVE CALIGOR, M.D., Clinical Professor of Psychiatry and Associate Director, Residency Training in Psychiatry, NYU School of Medicine; Chair, Psychodynamic Psychotherapy Division, and Training and Supervising Analyst, Columbia University Center for Psychoanalytic Training and Research.

ROBERT J. DERUBEIS, PH.D., Professor and Chair, Department of Psychology, School of Arts and Sciences, University of Pennsylvania.

ANDREW J. GERBER, M.D., PH.D., Assistant Professor of Clinical Psychiatry, Division of Child and Adolescent Psychiatry, New York State Psychiatric Institute.

ROBERT A. GLICK, M.D., Professor of Clinical Psychiatry, Columbia University College of Physicians and Surgeons; Past Director and Training and Supervising Analyst, Columbia University Center for Psychoanalytic Training and Research.

MARK J. HILSENROTH, PH.D., ABAP, Professor of Psychology, Derner Institute of Advanced Psychological Studies, Adelphi University.

ANDREW C. LEON, PH.D., Professor of Biostatistics in Psychiatry and Professor of Public Health, Weill Cornell Medical College.

KENNETH N. LEVY, PH.D., Associate Professor of Psychology, Pennsylvania State University.

RAYMOND A. LEVY, PSY.D., Clinical Director, Psychotherapy Research Program, Massachusetts General Hospital.

BARBARA MILROD, M.D., Professor of Psychiatry and Attending Psychiatrist, Weill Cornell Medical College Department of Psychiatry.

STEVEN P. ROOSE, M.D., Professor of Clinical Psychiatry, Columbia University College of Physicians and Surgeons; Research Psychiatrist, New York State Psychiatric Institute; Chair, Research Committee, Columbia Center for Psychoanalytic Training and Research.

BRET R. RUTHERFORD, M.D., Assistant Professor of Clinical Psychiatry, Columbia University College of Physicians and Surgeons; faculty, Columbia University Center for Psychoanalytic Training and Research.

KATHERINE SHEAR, M.D., Marion E. Kenworthy Professor of Psychiatry, Columbia University School of Social Work.

MICHAEL E. THASE, M.D., Professor of Psychiatry, University of Pennsylvania School of Medicine.

DREW WESTEN, PH.D., Professor, Department of Psychology and Department of Psychiatry and Behavioral Sciences, Emory University.

**APPENDIX B: THE PSYCHOANALYTIC OUTCOME
STUDY ADVISORY BOARD**

- ROBERT A. GLICK, M.D., Columbia University Center for Psychoanalytic Training and Research.
- STEVEN T. LEVY, M.D., Emory University Psychoanalytic Institute.
- ROBERT MICHELS, M.D., Columbia University Center for Psychoanalytic Training and Research.
- ERIC NUETZEL, M.D., Saint Louis Psychoanalytic Institute.
- STEVEN P. ROOSE, M.D., Columbia University Center for Psychoanalytic Training and Research.
- HARRIET WOLFE, M.D., San Francisco Center for Psychoanalysis.

REFERENCES

- ABLON, S.J., & JONES, E.E. (2005). On analytic process. *Journal of the American Psychoanalytic Association* 53:541–568.
- BARBER, J.P. (2009). Toward a working through of some core conflicts in psychotherapy research. *Psychotherapy Research* 19:1–12.
- CALIGOR, E., HILSENROTH, M.J., DEVLIN, M., RUTHERFORD, B.R., TERRY, M., & ROOSE, S.P. (2012). Will patients accept randomization to psychoanalysis? A feasibility study. *Journal of the American Psychoanalytic Association* 60:337–360.
- CALIGOR, E., STERN, B.L., HAMILTON, M., MACCORNACK, V., WININGER, L., SNEED, J., & ROOSE, S.P. (2009). Why we recommend analytic treatment for some patients and not for others. *Journal of the American Psychoanalytic Association* 57:677–694.
- CHERRY, S., WININGER, L., & ROOSE, S.P. (2009). A prospective study of career development and analytic practice: The first five years. *Journal of the American Psychoanalytic Association* 57:703–720.
- GERBER, A.J., KOCSIS, J.H., MILROD, B., ROOSE, S.P., BARBER J.P., THASE, M.E., PERKINS, P., & LEON, A.C. (2011). A quality-based review of randomized controlled trials of psychodynamic psychotherapy. *American Journal of Psychiatry* 168:19–28.
- GLICK, R., EAGLE, P., LUBER, B., & ROOSE, S.P. (1996). The fate of training cases. *International Journal of Psychoanalysis* 77:803–812.
- HILSENROTH, M., BLAGYS, M., ACKERMAN, S., BONGE, D., & BLAIS, M. (2005). Measuring psychodynamic-interpersonal and cognitive-behavioral techniques: Development of the Comparative Psychotherapy Process Scale. *Psychotherapy* 42:340–356.
- HOLLON, S.D., EVANS, M.D., AUERBACH, A., DERUBEIS, R.J., ELKIN, I., LOWERY, A., KRIS, M.R., GROVE, W.M., TUASON, V.B., & PIASECKI, J.M. (1988). Development of a system for rating therapies for depression:

- Differentiating cognitive therapy, interpersonal therapy, and clinical management pharmacotherapy. Unpublished manuscript.
- HUBER, D., & KLUG, G. (2005). Munich Psychotherapy Study (MPS): Preliminary results on process and outcome of psychoanalytic psychotherapy—a prospective psychotherapy study with depressed patients. *Psychotherapie Psychosomatik Medizinische Psychologie* 55:101.
- KERNBERG, O.F. (1999). Psychoanalysis, psychoanalytic psychotherapy and supportive psychotherapy: Contemporary controversies. *International Journal of Psychoanalysis* 80:1075–1091.
- LEON, A.C., SOLOMON, D.A., MUELLER, T.I., ENDICOTT, J., POSTERNAK, M., JUDD, L.L., SCHEFFLER, P., AKISKAL, H.S., & KELLER, M.B. (2000). A brief assessment of psychosocial functioning of subjects with bipolar I disorder: The LIFE-RIFT. *Journal of Nervous & Mental Disease* 188: 805–812.
- LEON, A.C., SOLOMON, D.A., MUELLER, T.I., TURVEY, C.L., ENDICOTT, J., & KELLER, M.B. (1999). The Range of Impaired Functioning Tool (LIFE-RIFT): A brief measure of functional impairment. *Psychological Medicine* 29:869–878.
- MILROD, B., LEON, A.C., BUSCH, F., RUDDEN, M., SCHWALBERG, M., CLARKIN, J., ARONSON, A., SINGER, M., TURCHIN, W., KLASS, E.T., GRAF, E., TERES, J.J., & SHEAR, M.K. (2007). A randomized controlled clinical trial of psychoanalytic psychotherapy for panic disorder. *American Journal of Psychiatry* 164:265–272.
- ROTHSTEIN, A. (1990). On beginning with a reluctant patient. In *On Beginning an Analysis*, ed. T. Jacobs & A. Rothstein. Madison, CT: International Universities Press, pp. 153–162.
- SHEDLER, J., & WESTEN, D. (2004a). Refining personality disorder diagnosis: Integrating science and practice. *American Journal of Psychiatry* 161:1350–1365.
- SHEDLER, J., & WESTEN, D. (2004b). Dimensions of personality pathology: An alternative to the five-factor model. *American Journal of Psychiatry* 161:1743–1754.
- SHEDLER, J., & WESTEN, D. (2007). The Shedler-Westen Assessment Procedure (SWAP): Making personality diagnosis clinically meaningful. *Journal of Personality Assessment* 89:41–55.
- STRUNK, D.R., DERUBEIS, R.J., CHIU, A.W., & ALVAREZ, J. (2007). Patients' competence in and performance of cognitive therapy skills: Relation to the reduction of relapse risk following treatment for depression. *Journal of Consulting & Clinical Psychology* 75:523–530.
- TRIVEDI, M.H., RUSH, A.J., WISNIEWSKI, S.R., NIERENBERG, A.A., WARDEN, D., RITZ, L., NORQUIST, G., HOWLAND, R.H., LEBOWITZ, B., MCGRATH, P.J.,

- SHORES-WILSON, K., BIGGS, M.M., BALASUBRAMANI, G.K., & FAVA, M. (2006). Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: Implications for clinical practice. *American Journal of Psychiatry* 163:28–40.
- VAUGHAN, S.C., MARSHALL, R., MACKINNON, R., & ROOSE, S.P. (1994). Current psychotherapy research methodology applied to psychoanalysis: A feasibility study. *Journal of Psychotherapy Practice & Research* 3: 334–340.
- VAUGHAN, S.C., SPITZER, R., DAVIES, M., & ROOSE, S. (1997). The definition and assessment of analytic process: Can analysts agree? *International Journal of Psychoanalysis* 78:959–973.
- WALLERSTEIN, R.S. (1994). Psychotherapy research and its implications for a theory of therapeutic change: A forty-year overview. *Psychoanalytic Study of the Child* 49:120–141.
- WARE, J.H., & HAMEL, M.B. (2011). Pragmatic trials: Guides to better patient care? *New England Journal of Medicine* 364:1685–1687.
- WESTEN, D., & SHEDLER, J. (2007). Personality diagnosis with the Shedler-Westen Assessment Procedure (SWAP): Integrating clinical and statistical measurement and prediction. *Journal of Abnormal Psychology* 116:810–822.

Steven P. Roose
1115 Park Avenue
New York, NY 10028
E-mail: spr2@columbia.edu