



Published in final edited form as:

Clin Trials. 2008 ; 5(3): 253–261. doi:10.1177/1740774508091965.

Addressing the challenges of a cross-national investigation: lessons from the Pittsburgh-Pisa study of treatment-relevant phenotypes of unipolar depression*

Ellen Frank^a, Giovanni B Cassano^b, Paola Rucci^a, Andrea Fagiolini^a, Luca Maggi^b, Helena C Kraemer^c, David J Kupfer^a, Bruce Pollock^d, Robert Bies^a, Vishwajit Nimgaonkar^a, Paul Pilkonis^a, M Katherine Shear^e, Wesley K Thompson^a, Victoria J Grochocinski^a, Paolo Scocco^f, Joan Buttenfield^a, and Rocco Nicola Forgiione^b

^aDepartment of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

^bDepartment of Psychiatry, Neurobiology, Pharmacology, Biotechnology, University of Pisa, Pisa, Italy

^cDepartment of Psychiatry and Behavioral Sciences, Stanford University, Palo Alto, CA, USA

^dDepartment of Psychiatry, University of Toronto, Toronto, Canada

^eSchool of Social Work, Columbia University, New York, NY, USA

^fDepartment of Mental Health, Padua, Italy

Abstract

Background—To date, no cross-national RCT has addressed the mechanisms underlying the relative success of pharmacological and psychotherapeutic interventions for depression. A multi-site clinical trial that includes psychotherapy as one of the treatments presents numerous challenges related to cross-site consistency and communication.

Purpose—This report describes how those challenges were met in the study “Depression: The Search for Treatment Relevant Phenotypes”, being carried out at the University of Pittsburgh and the University of Pisa, Italy.

Methods—Implementing the study required the investigators to address methodological and practical challenges related to the different requirements of the two Institutional Review Boards (IRBs), psychotherapy training, independent evaluator training, patient recruitment, development of common tools for data entry, quality control and generation of weekly reports of patient progress as well as establishing a similar clinical and research framework in two countries with substantially different health care systems.

Results—By having bilingual investigators and staff members who spent time at one another’s sites, making use of frequent conference-call staff meetings and being flexible within the bounds of the sometimes contradictory requirements of the IRBs, the investigators were able to meet the human subjects protection requirements of both institutions, surmount language barriers to

*The investigators of the study *Depression: The Search for Treatment-Relevant Phenotypes* are: Principal Investigator (PI): E. Frank (Pittsburgh), Co-PI: G.B. Cassano (Pisa), Pittsburgh Site Co-investigators: H. C. Kraemer; B. Pollock, M. K. Shear, P. Pilkonis, A. Fagiolini, P. Rucci, H. Swartz, D. Kupfer, R. Bies, I. Soreca; Pisa Co-Investigators: L. Dell’Osso, M. Mauri, S. Banti, A. Benvenuti, L. Maggi, M. Miniati, A. Papasogli, M. Sacttoni, C. Carmassi.

© Society for Clinical Trials 2008

Author for correspondence: Ellen Frank, Ph.D. Western Psychiatric Institute and Clinic, 3811 O’Hara Street, Pittsburgh, PA 15213. frank@upmc.edu.

consistent therapist and evaluator training and develop common tools for study management. As a result, recruitment goals were met at both sites and retention rates were high. One instance of inconsistent implementation of the protocol was corrected within the first year.

Limitations—This study was conducted in two Western cultures by researchers with long-standing collaboration. Our findings may not be generalizable to other countries or research settings.

Conclusions—The implementation of a cross-national protocol and the adoption and maintenance of common procedures is possible when investigators are aware of the challenges this may present and are proactive in trying to address them.

Background

In the last decade, international collaborative studies, including clinical trials that involved a large number of countries and enrolled thousands of patients in different continents have been undertaken in many areas of medicine, particularly in oncology [1,2] and cardiology [3–5]. International clinical trials have been conducted in psychiatry as well, typically with a focus on the comparative efficacy of drug treatments (for instance clozapine vs olanzapine to reduce suicidality [6] or alprazolam vs imipramine vs placebo to reduce panic disorder symptomatology [7]). Only two cross-national trials compared non-pharmacologic interventions: one is an ongoing trial on treatments for specific phobia in children in the US and Sweden (NIMH NCT00051220) while the second was designed to compare delivery methods for cognitive-behavioral therapy in panic disorder in Scotland and Australia [8]. No cross-national randomized clinical trial has been conducted to date to identify the mechanisms underlying the relative success of pharmacological and psychotherapeutic interventions for depression. This is a notable gap in the literature, because there is no evidence that these mechanisms are the same across different cultural and health care settings. In fact meta-analyses, including trials from different countries, fail to capture this information because they are fraught with difficulties resulting from differences in diagnostic and assessment methods, in addition to the technical problems related to how best to assess statistical significance. Answering questions regarding *who* benefits from specific treatments and *how* this occurs requires comparisons across large diagnostically and ethnically homogeneous data sets.

In an attempt to fill this gap, investigators at the University of Pittsburgh (USA) and the University of Pisa (Italy), in the context of an on-going collaborative project to develop a dimensional assessment of mood and anxiety psychopathology, decided to conduct a two-site, cross-national clinical trial aimed at identifying demographic, biological, and psychopathological mediators and moderators of treatment response with psychotherapy, antidepressant, or their combination. According to Kraemer *et al.* [9], *moderators* identify on whom and under what circumstances treatments work. Moderators are baseline characteristics of subjects that are uncorrelated with treatment and can be shown to have an interactive effect with treatment on outcome. *Mediators* differ from moderators in that they characterize changes or events that occur during treatment that are correlated with treatment and identify possible mechanisms through which a treatment might achieve its effects. The main empirical questions to be answered in the depression phenotypes study were: (1) what are the mediators and moderators of depression treatment outcomes, and (2) whether they differ between cultures. While collaboration between the Pittsburgh and Pisa sites on measure development had proven a relatively simple task [10], mounting a treatment study involving both psychotherapeutic and pharmacologic interventions across differing languages, health systems, and research cultures proved considerably more challenging.

In the present report we describe the rationale for our study of treatment-relevant depression phenotypes, the challenges faced in initiating and carrying out a complex cross-national study protocol and how we addressed those challenges.

Rationale for the depression phenotypes study

Despite decades of clinical trial experience in major depression, our understanding of how best to achieve durable recovery is quite limited. Response rates range between 65 and 75% [11,12]. Full remission is seen in only one in three subjects (20–45% [13–15], 28% [12]). Those who achieve remission remain at considerable risk of relapse (50% [16]; 60–80% [17]). Even those who recover symptomatically generally fail to show a return to pre-morbid levels of functioning [18]. Meta-analyses comparing percentages of full remission in patients treated with pharmacotherapy or psychotherapy do not provide evidence for the superiority of one treatment over the other [15].

In designing the present study, our interest was in identifying those subgroups of patients suffering from unipolar depression that respond best to antidepressant medication vs depression-specific psychotherapy or to sequences of those two treatments, achieve full remission of symptoms and return of functioning, and are able to sustain their recovery and improved functioning through an extended well interval. Previous studies directly comparing the efficacy of drugs and psychotherapy attempted to identify *post hoc* those factors that were associated with response to one or the other treatment [19–21]. In this study, we made *a priori* predictions about the mechanisms by which remission is achieved and sustained in each form of treatment. Thus, we were able to articulate hypotheses about those factors that might mediate and moderate medication vs psychotherapy response. We used random assignment to the initial treatment as the best defense against selection bias and to ensure that these factors were equally balanced between the two treatment groups. The ultimate goal of the investigation is to produce clinically useful algorithms for deciding whether an individual outpatient with moderately severe depression should be treated with medication, psychotherapy, or the combination and to determine whether the same algorithms hold in two different cultures.

With these objectives in mind, we designed a multifaceted study conducted within the framework of an acute and continuation treatment protocol involving pharmacotherapy, psychotherapy, and their combination.

Study aims

Primary aims of the study are to examine: (1) the impact of the type and number of *lifetime* mood disorder features a patient endorses on the outcome of treatment of major depression; (2) the impact of the type and number of *lifetime* anxiety features a patient endorses on the outcome of treatment of major depression; (3) the impact of treatment exposure (conceptualized as specificity of the psychotherapy provided and/or consistency of pharmacotherapy exposure) on the outcome of treatment of major depression. Secondary aims, hypothesis-generating rather than hypothesis-testing, are to determine (1) which subgroups of depressed patients, characterized using the above variables, show optimal functional outcomes and the most complete and sustained resolution of symptoms when exposed to psychotherapy alone, pharmacotherapy alone, which subgroups require a treatment sequence involving the augmentation of psychotherapy with pharmacotherapy or pharmacotherapy with psychotherapy, and which subgroups do not achieve remission with any of these interventions; (2) which set of constructs (DSM-IV Axis II personality pathology, features of temperament or the mood and anxiety spectrum we have defined) is most robust in predicting time to stabilization and residual functional impairment among

patients with major depression; and (3) whether particular genetic polymorphisms (such as variants of the serotonin transporter, *SLC6A4*), are correlated with treatment response.

We hypothesized that the endorsement of spectrum features and their impact on treatment response would differ between sites as a result of the reporting style in the two cultures and the expectations that patients have in the US insurance-based health care system compared with the Italian national health care system that covers almost entirely the costs of care for all citizens. In addition, we hypothesized that the pharmacokinetic of the study drug and therefore the speed of response and tolerance would differ between sites as a result of the dietary habits and the body mass index in the US and in the Italian samples.

Overview of study protocol

The platform for this investigation is a treatment protocol in which 290 subjects with a non-psychotic major depressive episode are randomly assigned to pharmacotherapeutic or psychotherapeutic intervention for depression. The pharmacotherapy we chose to study, escitalopram, is a member of the selective serotonin reuptake inhibitor (SSRI) class of antidepressants. Interpersonal psychotherapy (IPT [22,23]), an empirically-validated manual-based treatment for major depression was selected as the psychotherapeutic intervention. If the initial treatment is not successful in bringing about remission, subjects receive augmentation with the other treatment (pharmacotherapy or psychotherapy). See Figure 1.

Measures

The primary outcome measures of the trial are the participants' weekly scores on the Hamilton Rating Scale for Depression (HRSD) [24]. These scores are used to determine change in depressive symptoms over time, and are crucial for determining remission of depression (in the acute treatment phase) and depressive relapse (in the continuation treatment phase). The weekly HRSD scores also constitute the basis for making decisions about adding/changing treatments in the protocol.

The other core study assessments are the four self-report mood and anxiety 'spectrum' measures, that were developed in parallel in English and Italian by clinical investigators, including many of the authors of the present study, with the aim of exploring the broad range of mood and anxiety symptoms that surround the DSM-IV defined disorders.

Developing and validating these measures in a cross-cultural context required ten years of work on the part of the investigators. The challenge was to determine a minimum common denominator of relevant questions to explore spectrum phenomenology in both cultures, using simple language, and avoiding jargon or idiomatic expressions. The term 'spectrum' is used [25] to capture the continuum between the core symptoms of each disorder and associated prodromal, atypical, and subclinical psychopathological expressions. Thus, this continuum includes (i) typical and atypical symptoms of the primary DSM-IV disorder, (ii) behavioral patterns and other features related to the core symptoms that may either be prodromal, precursor states, or sequelae of a previously expressed disorder, and (iii) temperamental or personality traits. The spectrum instruments were developed in two versions: a *lifetime* version, exploring symptoms and experiences occurring over the entire life of an individual and a *last-month* version, exploring symptoms that occurred in the month prior to the assessment.

Other study assessments include established interview-based and self-report assessments of mood psychopathology as well as measures of temperament and personality, measures of somatic symptoms/collateral effects of pharmacotherapy, and measures of functional

impairment and quality of life. Assessment instruments that had been validated in both English and Italian were chosen for the study. However, a careful review of each instrument by bilingual investigators was carried out to ensure that items were worded in the same way in both languages.

Two indirect measures of adherence to the pharmacotherapy regimen or pharmacotherapy exposure were used at both study sites: population pharmacokinetics and pill count [26]. Psychotherapy exposure was assessed using methods developed by the first author for the evaluation of audiotapes of interpersonal psychotherapy sessions [27,28]. Earlier studies demonstrated that individual patient/therapist dyads vary on the intensity of interpersonal focus, or IPT 'specificity,' and that this variable is significantly associated with treatment outcome [27,29]. Finally, subjects were given the option of whether or not to participate in pharmacogenetic studies. The focus of these studies includes candidate polymorphisms relevant to monoamine function and the mechanism of antidepressant response. Blood samples were collected at both sites using a common protocol, to enable extraction of genomic DNA.

Analytic plan

The data-analytic approach rests upon the concepts and definitions of moderators and mediators of outcome in the framework of randomized clinical trials [9]. This data analytic approach is implemented via an application of hierarchical linear models (HLMs [32]) for the primary outcomes, i.e., change in weekly HDRS scores (slope). Treatment group (IPT/SSRI) and lifetime spectrum score are the main effects in this model. Moderator analyses proceed by including the interaction of the lifetime spectrum score with treatment group. This interactive effect is examined as potential evidence that the spectrum measures moderate treatment effect, even in the absence of a main effect [9]. When a main effect of the spectrum measure is present but no interactive effect with treatment group is found, we would conclude that the spectrum measure is a non-specific predictor of treatment outcome. Site is also included in the HLMs as a potential moderator of treatment effect, by interacting treatment group with site and slope. Cox proportional-hazard survival regression analyses are used to identify clinical moderators of the secondary outcomes (time to remission and time to relapse) by examining interactions of treatment group and spectrum measures similar to the above. Finally, HLMs are also used to test the hypothesis that the change in spectrum measures mediates the effect of treatment on residual functional impairment, by testing the interactions of change in spectrum scores with treatment group in HLMs with residual functional impairment as outcomes.

The secondary goals of the investigation are represented in the hypothesis that the lifetime spectrum constructs can be used along with traditional predictors, such as baseline severity of depression to identify which patients subgroups are likely to remit with IPT alone, or with SSRI alone and which are likely to require augmentation treatment to achieve remission. Signal detection analysis, a variation of recursive partitioning of patients based on classification trees [33] was used for identifying interactions and subgroups and for providing insight into the characteristics of patients likely to remit after receiving each of the two treatments, with and without augmentation.

Power analysis

We conservatively computed power based on a linear regression model with change in HDRS score from baseline as the dependent variable and treatment group and lifetime spectrum score (dichotomized at the median) entered as independent variables and interacted with one another. Sample size and power were calculated to detect an interaction effect in this linear model. This is a conservative model for power calculations because it ignores the

intermediate HRSD scores between baseline and end of phase (acute or continuation), and because we will actually employ the continuous spectrum scores in the HLMs, not median splits. Assuming equal cell sizes (which is reasonable because baseline variables are independent of treatment assignment), the power to detect the interaction term (say γ) is a function of γ , σ and N , where σ is the error standard deviation and N is the total sample size. With $N = 290$ and σ estimated at 6.5 (obtained from prior depression studies utilizing longitudinal HRSD scores and validated by the error standard deviation estimates obtained from the current study), we are able to detect a $\gamma = \pm 3$ or larger with >95% power at level $\alpha = 0.05$. This corresponds to a three point difference or higher in HRSD change with SSRI compared to IPT between low spectrum and high spectrum groups. These power calculations were performed using the software package Power and Sample Size Calculation, version 2.1.31 [34].

Challenges

Informed consent and genetic study

To comply with National Institute of Mental Health requirements, the Pisa site had to obtain the Federalwide assurance. As this procedure was completely new to the Pisa site, the active involvement of the Rector and President of the Bioethics Committee of the University of Pisa was needed in order to obtain their approval. All study personnel at both sites needed to be trained and certified on the National Institutes of Health Human Participant Protection modules regarding privacy and patient safety which are available only in English.

The University of Pittsburgh IRB and the University of Pisa Ethics Committee have very different and completely inconsistent requirements for consent forms. At Pisa, there is a specific limit of five pages on the length of consent forms. In contrast, at Pittsburgh, the IRB requires a level of detail that means that an approvable consent form for a study such as this exceeds 20 pages. Therefore, it was necessary to obtain the Pittsburgh IRB's agreement on the minimum set of the key elements to include in the Pisa consent in order to have consistency between sites. Furthermore, the Pittsburgh IRB had to approve the Ethics Committee protocol and consent forms for the Pisa site, all of which therefore had to be translated into English for their approval. Difficulties were encountered in obtaining the approval for the genetic study in Italy. There are strict norms that govern privacy in Italy and affect genetic analyses, shipment of the genetic materials to other laboratories and storage of genetic samples. This delayed the collection of genetic samples by two years at the Pisa site. Fortunately, we were able to obtain DNA from several of the Pisa participants who had completed the study during those first two years.

Drug treatment

When the study was designed, the proposed study medication, escitalopram, was not yet on the market in Italy. Therefore, a one-year pilot study was conducted using citalopram. When the full study began, pharmacotherapists at Pisa started using escitalopram without having had any previous experience with the drug and its side effects. The advice of Pittsburgh colleagues proved to be crucial in suggesting how to adjust the time of administration and how to manage side effects. However, even when consensus was achieved on clinical management, escitalopram appeared to be tolerated differently at the two sites, with Pisa subjects experiencing noticeably more distressing side effects. Many hypotheses were formulated as to why this was occurring. One hypothesis was that the pills administered at Pisa and Pittsburgh contained a different percentage of active compound, the second was that for some reason the same dose of escitalopram was leading to higher blood concentrations at the Pisa site. Both of these potential explanations were tested empirically

and disproven. In the end, the protocol was modified at both sites to allow more flexibility in the dosage titration or reduction as a consequence of side effects.

Primary outcomes

The primary outcomes of the study are the weekly HRSD scores. These are used to determine change in depressive symptoms from baseline over the course of the acute phase and change in depressive symptoms from time of remission over the course of the maintenance phase. Three secondary outcomes are also defined based on weekly HRSD scores: response, defined by a 50% decrease in the baseline scores on the HRSD, remission, defined by three consecutive weeks during which the HRSD score averages ≤ 7 and relapse, defined by a substantial clinical worsening evaluated twice within a seven-day period, by the treating clinician and then confirmed by an independent clinical psychiatrist. Relapse was confirmed if both evaluations were consistent with a diagnosis of a Major Depressive Episode and a HRSD score of ≥ 15 . An additional secondary outcome of interest was the change in functional impairment from baseline, determined by using the score on a self-report instrument, the Work and Social Adjustment Scale (WSAS [30]).

Training of personnel

This multi-faceted study, directed by an American clinical psychologist (EF) and an Italian psychiatrist (GBC) has required the expertise and skills of a cross-national multi-disciplinary research group, including, at the Pittsburgh site, two part-time pharmacotherapists, five part-time psychotherapists, one data manager and three administrative staff; at the Pisa site, six part-time psychiatrists (four psychotherapists and two pharmacotherapists), two psychologists (independent evaluators), one statistician, one data manager, one biologist, and two full-time administrative staff. The PI and co-PIs at both sites provided constant oversight of the project.

Investigators at the two sites had been collaborating on instrument development projects for more than five years at the time the depression phenotypes study was designed. The PIs at both sites and several of the investigators at Pisa were bilingual in English and Italian. The project coordinator at the Pisa site (PR) had previously worked at Pittsburgh for more than a year and one of the pharmacotherapists at the Pittsburgh site had done some of his training at Pisa.

The five psychotherapists at Pittsburgh are mental health professionals from a variety of backgrounds including clinical and counseling psychology, social work, and psychiatric nursing, all of whom had had extensive experience in the conduct of IPT at the outset of the study. In contrast, because individuals with such backgrounds are not ordinarily qualified to perform psychotherapy in Italy, the four clinicians selected to participate as psychotherapists at the Pisa site were all psychiatrists. Only two of the Pisa psychotherapists had had any previous psychotherapy training and all four had a strong biological orientation. The methods involved in their training have been described in detail elsewhere [31]. Briefly, the Pisa therapists were initially trained in IPT during an intensive didactic seminar conducted at Pisa by PS, who works in Padua, Italy and at the time was the only Italian psychiatrist experienced in interpersonal psychotherapy. PS had, in turn, received some of his original training in IPT through a seminar conducted in Italy several years earlier by the first author. Prior to the course, the trainees were given the Italian translation of the IPT manual. They were asked to read it carefully and start practicing IPT-like assessments and case-conceptualization with their pharmacotherapy patients. Following the didactic course, each of the Pisa trainees received supervision by phone of 12 IPT sessions with each of two cases by PS. The first author met with the Italian supervisor and trainees approximately every four months for the first three years of the study to review cases and discuss issues in the

implementation of IPT. Thus, the IPT practiced at Pisa tends to mirror the IPT practiced in the first author's research group at Pittsburgh particularly closely.

Quality control of intervention and assessments

Consistency across sites in the administration and rating of the primary outcome measure is a crucial issue. A bilingual psychiatrist from Pisa (CC) was trained over a period of one year at Pittsburgh in the administration of the HRSD and was certified as the 'gold standard' rater for the Pisa site. Her certification was obtained when the discrepancy with the Pittsburgh gold standard rater (JB) did not exceed three points on the total score or one point on single items in five consecutive ratings. Two psychologists were then trained as raters at Pisa by CC. They co-rated 15 HRSD interviews with CC and were certified when they reached the level of agreement described above.

IPT sessions were audiotaped at both sites and then rated using the Therapy Rating Scale [28]. This scale allows raters to assess the adherence of therapists to the IPT model by discriminating techniques typical of this therapy from those used in cognitive therapy and psycho dynamic psychotherapy, the two most common other approaches to the psychotherapeutic treatment of depression.

The Therapy Rating Scale was first translated into Italian by PS and PR. It was then back-translated into English by the first author. All three then met to resolve any discrepancies that arose in the back translation. When they were satisfied that all items had the same meaning in Italian and English, two Italian psychologists were trained to rate IPT sessions. When they had achieved inter-rater reliability of 0.85 on each individual item, they were certified as study raters. The maintenance of this level of inter-rater reliability has been tested throughout the study by having the raters co-rate 10% of the therapy sessions.

Patient recruitment

Within the Italian health care system, access to psychiatric facilities is usually possible by referral from primary care physicians. However, self-referral to psychiatric services is also permitted. Different from the US, Italian law does not allow any announcements regarding clinical studies in public places or in the media, nor does it allow the reimbursement of patients for their time and expenses related to participation in the study. Therefore, patient recruitment must be based almost exclusively on the natural flow of patients seeking treatment at the outpatient unit of the psychiatric clinic. Strategies to increase recruitment are limited and consist mainly in trying to increase referrals from primary care physicians. In contrast, in the US it is possible to use public service announcements, as well as paid advertisements of the availability of the study on radio and television, all of which were used at the Pittsburgh site. The major incentive for patients at Pittsburgh was the availability of ≈ 10 months of treatment for their depression at no cost, including no charge for medication. Psychotropic drugs included in class A of the Italian National Formulary (INF), such as antidepressants, are free of charge, while in the US these medications can be very costly. Thus, for US patients, a study that pays medication costs is very attractive. In both Italy and the US, psychotherapies are generally at least partially paid by patients; however, in Italy costs for patients are minimal, compared with those incurred in the US, especially by the uninsured. However, the study did represent the first time that IPT was available at the Pisa site and this provided an incentive for some of the subjects recruited there. Despite these differences in regulations and incentives, patient flow during the first four years of the study has been virtually identical at the two sites.

Trial communications, data management, and study monitoring

Frequent contacts were held via weekly scheduled conference calls and site visits between the investigators and the staff at the two sites throughout the five years of the study. During the conference calls, the Italian and US team reviewed progress in patient recruitment and occurrence of severe and non-severe adverse events, as well as clinical progress of patients based on a detailed clinical report on each active patients generated by computer.

These calls were an important opportunity to prevent the occurrence of differences between sites resulting from a misinterpretation of the protocol.

One example is that at Pisa site the occurrence of the relapse was mistakenly deemed a reason for terminating the patient from the protocol while at Pittsburgh site patients continued to be followed up until the end of study, even in the event of a relapse. This misinterpretation was identified and corrected in the first year of the study.

The timely production of automatic reports required prompt data entry and data merging. To this end, the database management framework provided by the Pittsburgh center was translated into Italian to make the data-entry interface user-friendly. The data quality was checked on regular basis, ensuring that data analysis could be promptly conducted. In order to ensure that all human subjects protections were being adhered to and that all study records were complete from an Institutional Review Board perspective, completeness and accuracy of patient files and other relevant records of all projects at the Pittsburgh site were monitored on a regular basis. We took advantage of this monitoring tradition to further increase cross-site communication by having the Program Coordinator at each site travel to the other site annually to carry out study monitoring there.

Conclusions

The implementation of a cross-national protocol and the adoption of identical procedures and assessments in two countries have the potential to identify genuine differences (or similarities) in terms of remission rates and mediators/moderators of depression treatment outcomes that meta-analyses, even those including trials from different countries, fail to capture. Given the extent to which the developed world is becoming a 'global village' when it comes to the practice of psychiatry and the extent to which published results are assumed to apply throughout the developed world, it seems essential that we demonstrate that this is so. Such demonstrations are possible only through multisite, cross-national studies. Implementing such studies in two countries using a common language is difficult enough. We have described here some of the challenges encountered (and, we believe, largely met) when we attempted to mount such a study in countries using different languages, having at least moderately different cultures, and very different health systems.

Among the factors leading to the relatively smooth mounting of this cross-national study were: (1) the pre-existing collaborative relationship between the two PIs; (2) more than five years of prior collaboration among almost all the investigators involved in the study; (3) the fact that several of the Pittsburgh investigators had spent time at the Pisa site and vice versa; (4) the fact that there were at least two bilingual investigators at each site; (5) the willingness of the two IRBs to be flexible and respectful of each other's requirements and that, at both sites, the PIs had invested time and energy in developing working relationships with the IRB administrators and directors; and 6) the frequent contacts via regularly-scheduled conference calls and site visits between the investigators and the staff at the two sites. The last factor proved to be crucial in ensuring that recruitment goals were met at both sites and retention rates were high (79% during the acute treatment phase; 68% including the six-month continuation phase). It may be possible to conduct successful cross-national trials

without a prior history of collaboration, bilingual investigators, personal knowledge of how research is conducted at the other site, regular and frequent contact between the sites etc., but our experience would suggest that the more of these factors that are present, the more likely the success of implementation.

Acknowledgments

This work was supported by National Institute of Mental Health grants MH65376 (Drs. Frank and Cassano) and MH30915 (Dr. Frank), an investigator-initiated grant from Forest Research Institute (Dr. Frank) and Fondazione IDEA (Dr. Cassano).

References

1. Anak O, Van Cutsem E, Nordlinger B. The EORTC Gastrointestinal Tract Cancer Group: 40 years of research contributing to improved GI cancer management. *Eur J Cancer*. 2002; 38:S65–S70. [PubMed: 11858968]
2. Moy B, Goss PE. TEACH: Tykerb evaluation after chemotherapy. *Clin Breast Cancer*. 2007; 7:489–492. [PubMed: 17386127]
3. Yusuf S, Hawken S, Ounpuu S, et al. INTERHEART Study Investigators. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. *Lancet*. 2004; 364:937–952. [PubMed: 15364185]
4. Alexander JH, Reynolds HR, Stebbins AL, et al. TRIUMPH Investigators. Effect of tilarginine acetate in patients with acute myocardial infarction and cardiogenic shock: the TRIUMPH randomized controlled trial. *JAMA*. 2007; 297:1657–1666. [PubMed: 17387132]
5. Armstrong PW, Granger CB, Adams PX, et al. APEX AMI Investigators. Pexelizumab for acute ST-elevation myocardial infarction in patients undergoing primary percutaneous coronary intervention: a randomized controlled trial. *JAMA*. 2007; 3:43–51. [PubMed: 17200474]
6. Meltzer HY, Alphas L, Green AI, et al. International Suicide Prevention Trial Study Group. Clozapine treatment for suicidality in schizophrenia: International Suicide Prevention Trial (InterSePT). *Arch Gen Psychiatry*. 2003; 60:82–91. Erratum in: *Arch Gen Psychiatry* 2003; 60: 735. [PubMed: 12511175]
7. Buller R, Winter P, Amering M, et al. Center differences and cross-national invariance in help-seeking for panic disorder. A report from the cross-national collaborative panic study. *Soc Psychiatry Psychiatr Epidemiol*. 1992; 27:135–141. [PubMed: 1621138]
8. Kenardy JA, Dow MGT, Johnston DW, et al. A comparison of delivery methods of cognitive behavioral therapy for panic disorder: an international multi-center trial. *J Consult Clin Psychol*. 2003; 71:1068–1075. [PubMed: 14622082]
9. Kraemer HC, Wilson GT, Fairburn CG, Agras WS. Mediators and moderators of treatment effects in randomized clinical trials. *Arch Gen Psychiatry*. 2002; 59:877–883. [PubMed: 12365874]
10. Frank E, Shear MK, Rucci P, et al. Cross-cultural validity of the structured clinical interview for panic-agoraphobic spectrum. *Social Psychiatry and Psychiatric Epidemiology*. 2005; 4:283–291. [PubMed: 15834779]
11. Reifler BV. Play it again, Sam—depression is recurring. *N Engl J Med*. 2006; 354:1189–1190. [PubMed: 16540621]
12. Trivedi MH, Rush AJ, Wisniewski SR, et al. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. *Am J Psychiatry*. 2006; 163:28–40. [PubMed: 16390886]
13. Davidson JR, Meoni P, Haudiquet V, et al. Achieving remission with venlafaxine and fluoxetine in major depression: its relationship to anxiety symptoms. *Depress Anxiety*. 2002; 16:4–13. [PubMed: 12203668]
14. Entsuah AR, Huang H, Thase ME. Response and remission rates in different subpopulations with major depressive disorder administered venlafaxine, selective serotonin reuptake inhibitors or placebo. *J Clin Psychiatry*. 2001; 62:869–977. [PubMed: 11775046]

15. Casacalenda N, Perry JC, Looper K. Remission in major depressive disorder: a comparison of pharmacotherapy, psychotherapy and control conditions. *Am J Psychiatry*. 2002; 159:1354–1360. [PubMed: 12153828]
16. Perahia DG, Gilaberte I, Wang F, et al. Duloxetine in the prevention of relapse of major depressive disorder: double-blind placebo-controlled study. *Br J Psychiatry*. 2006; 188:346–353. [PubMed: 16582061]
17. Frank E, Kupfer DJ, Perel JM, et al. Three-year outcomes for maintenance therapies in recurrent depression. *Arch Gen Psychiatry*. 1990; 47:1093–1099. [PubMed: 2244793]
18. Judd LL, Akiskal HS, Zeller PJ, et al. Psychosocial disability during the long-term course of unipolar major depressive disorder. *Arch Gen Psychiatry*. 2000; 57:375–380. [PubMed: 10768699]
19. Elkin I, Gibbons RD, Shea MT, et al. Initial severity and differential treatment outcome in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *J Consult Clin Psychol*. 1995; 63:841–847. [PubMed: 7593878]
20. Sotsky SM, Glass DR, Shea MT, et al. Patient predictors of response to psychotherapy and pharmacotherapy: Findings in the NIMH Treatment of Depression Collaborative Research Program. *Am J Psychiatry*. 1991; 148:997–1008. [PubMed: 1853989]
21. Scott AIF, Freeman CPL. Edinburgh Primary Care Depression Study: Treatment outcome, patient satisfaction, and cost after 16 weeks. *Br Med J*. 1992; 304:883–997. [PubMed: 1392754]
22. Klerman GL, Weissmann MM. Interpersonal psychotherapy (IPT) and drugs in the treatment of depression. *Pharmacopsychiatry*. 1987; 20:3–7. [PubMed: 3547419]
23. Weissman MM, Markowitz JC. Interpersonal psychotherapy. Current status. *Arch Gen Psychiatry*. 1994; 51:599–606. [PubMed: 8042909]
24. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry*. 1960; 23:56–62. [PubMed: 14399272]
25. Cassano GB, Michellini S, Shear MK, et al. The panic-agoraphobic spectrum: a descriptive approach to the assessment and treatment of subtle symptoms. *Am J Psychiatry*. 1997; 154:27–38. [PubMed: 9167542]
26. Bies RR, Mulsant BH, Rosen J, et al. Population pharmacokinetics as a method to detect variable drug exposure in patients suffering from dementia with behavioral disturbances. *Am J Geriatric Pharmacother*. 2005; 3:87–91.
27. Frank E, Kupfer DJ, Wagner EF, et al. Efficacy of interpersonal psychotherapy as a maintenance treatment of recurrent depression. Contributing factors. *Arch Gen Psychiatry*. 1991; 48:1053–1059. Erratum in: *Arch Gen Psychiatry* 1992; 49: 401. [PubMed: 1845438]
28. Wagner EF, Frank E, Steiner S. Discriminating maintenance treatments for recurrent depression: Development and implementation of a rating scale. *Journal of Psychotherapy Research and Practice*. 1992; 1:280–290.
29. Spanier C, Frank E, McEachran AB, et al. The prophylaxis of depressive episodes in recurrent depression following discontinuation of drug therapy: integrating psychological and biological factors. *Psychol Med*. 1996; 26:461–475. [PubMed: 8733205]
30. Mundt JC, Marks IM, Shear MK, Greist JH. The work and social adjustment scale: a simple measure of impairment in functioning. *Br J Psychiatry*. 2002; 180:461–464. [PubMed: 11983645]
31. Frank E, Scocco P. IPT training and supervision in Italy: a model for the clinical psychiatrist across geographic and linguistic boundaries. *Bulletin of the International Society for Interpersonal Psychotherapy*. 2003; 2:2–4.
32. Gibbons RD, Hedeker D, Elkin I, et al. Some conceptual and statistical issues in analysis of longitudinal psychiatric data: Application to the NIMH Treatment of Depression Collaborative Research Program Dataset. *Arch Gen Psychiatry*. 1993; 50:739–750. [PubMed: 8357299]
33. Kiernan M, Kraemer HC, Winkleby MA, et al. Do logistic regression and signal detection identify different subgroups at risk?: Implications for the design of tailored interventions. *Psychological Methods*. 2001; 6:35–48. [PubMed: 11285811]
34. Dupont WD, Plummer WD. Power and sample size calculations: a review and computer program. *Controlled Clinical Trials*. 1990; 11:116–128. [PubMed: 2161310]

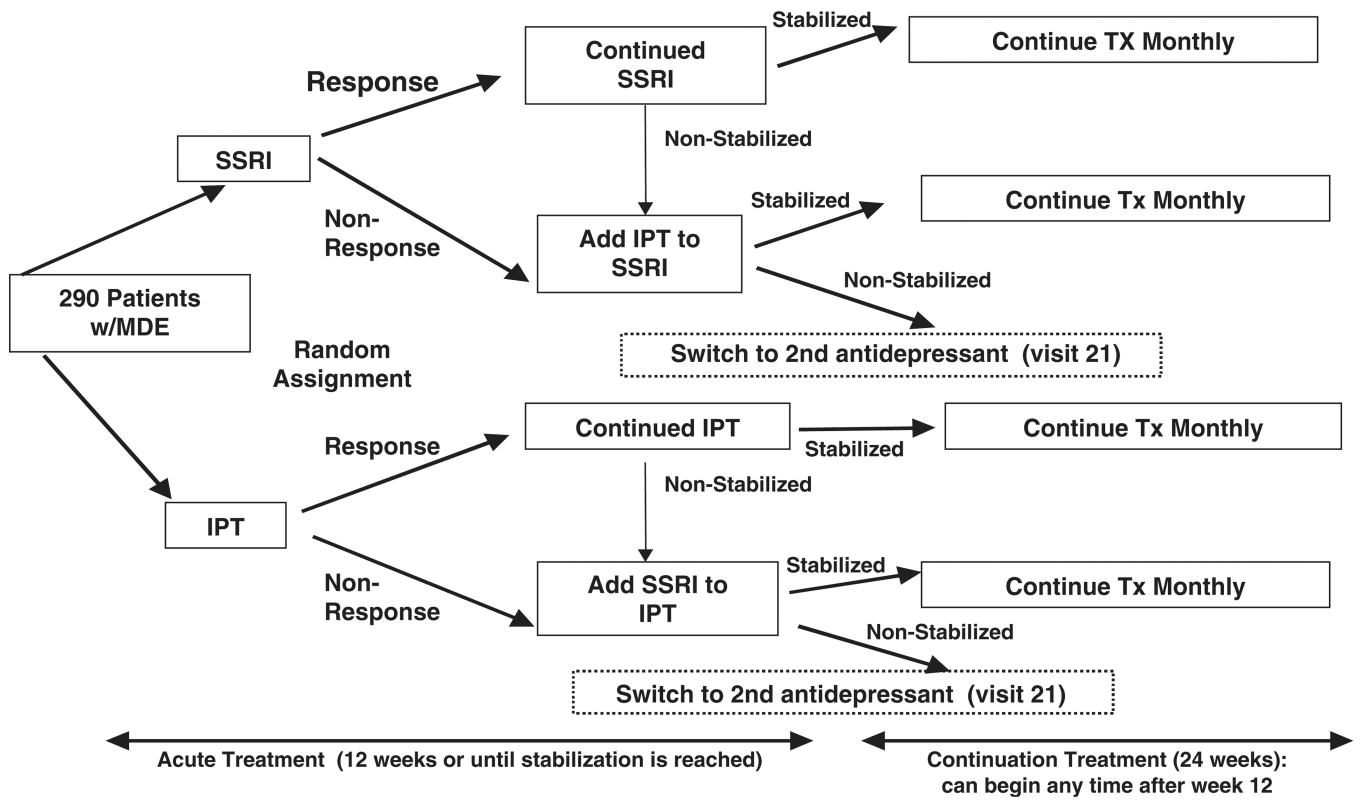


Figure 1.
Acute and continuation treatment protocol