ELSEVIER

Design Paper

Design of Department of Veterans Affairs Cooperative Study No. 420: Group Treatment of Posttraumatic Stress Disorder

Paula P. Schnurr, PhD, Matthew J. Friedman, MD, PhD, Philip W. Lavori, PhD, and Frank Y. Hsieh, PhD

VA National Center for PTSD, White River Junction, Vermont (P.P.S., M.J.F.); Dartmouth Medical School, Hanover, New Hampshire (P.P.S., M.J.F.); VA Cooperative Studies Program, Menlo Park, California (P.W.L., F.Y.H.); and Stanford University, Palo Alto, California (P.W.L.)

ABSTRACT: Posttraumatic stress disorder (PTSD) is a significant problem for a large number of veterans who receive treatment from the Department of Veterans Affairs (VA) health-care system. VA Cooperative Study 420 is a randomized clinical trial of group psychotherapy for treating PTSD among veterans who sought VA care. Participants (n = 360) at ten sites were randomly assigned to receive one of the two treatments: active treatment that embedded exposure therapy in a group context or comparison treatment that avoided trauma focus and instead addressed current interpersonal problems. Treatment was delivered weekly to groups of six participants for 30 weeks, followed by five monthly booster sessions. Follow-up assessments were conducted at the end of treatment (7 months) and the end of boosters (12 months) for all participants. Long-term follow-up data were collected for a subset of participants at 18 and 24 months. The primary outcome is PTSD severity; other symptoms, functional status, quality of life, physical health, and service utilization also were assessed. Data analysis will account for the clustering introduced by the group nature of the intervention. The pivotal comparison was at the end of treatment. Analyses of subsequent outcomes will concentrate on the question of the durability of effects. The study provides an example of how to address the unique challenges posed by multisite trials of group psychotherapy through attention to methodological and statistical issues. This article discusses these challenges and describes the design and methods of the study. *Control Clin Trials* 2001;22:74–88 © Elsevier Science Inc. 2001

KEY WORDS: *Psychotherapy research, posttraumatic stress disorder (PTSD), group therapy, military veterans*

Address reprint requests to: Paula P. Schnurr, PhD, National Center for PTSD (116D), VA Medical & Regional Office Center, 215 N. Main Street, White River Junction, VT 05009 (E-mail: Paula.Schnurr@Dartmouth.edu).

Received February 11, 2000; accepted September 18, 2000.

Controlled Clinical Trials 22:74–88 (2001)
© Elsevier Science Inc. 2001
655 Avenue of the Americas, New York, NY 10010

0197-2456/01/$–see front matter
PII S0197-2456(00)00118-5

course of the disease, regression towards the mean (sick patients improve without treatment, less sick patients are more likely to deteriorate), run of luck, bias, and placebo effect are also discussed. In Chapter 6 techniques of patient selection, randomization, choice of a control group, the principles of masking (blinding), as well as stopping rules and assessment of results are discussed in a very easy to read manner, with interesting and illuminating examples. The authors emphasize that "the randomized clinical trial is the gold standard for the assessment of any treatment, not only pharmacotherapy, but also surgical interventions, physiological therapy, dietary treatment, different types of nursing care, and preventive measures in the general population." Statistical analysis of randomized clinical trials as well as the importance of meta-analysis are explained in a manner interesting to the average clinician. The authors affirm that decisions taken in everyday clinical work must to the greatest extent be based on the results of clinical research. I would hope that clinical trial investigators take heed of their recommendation that patients who participated in a clinical trial be informed afterwards of which treatment they received and what the trial results were. Chapter 7 focuses on ethical issues related to the clinical decision making process using examples from everyday clinical practice. Issues of informed consent, the international recommendations from the Helsinki declaration and other ethical topics related to clinical research are also discussed. Chapter 8 affords the authors an opportunity to outline some of the common epidemiological and biostatistical approaches encountered in the medical literature. The reader is also provided with useful advice for evaluating medical journal articles.

Overall, the authors highlight the translation of research findings to the individual patient. The reader should experience an exciting, historical and up to date review of the evolution of medical practice, the principle of clinical decision making from diagnosis to treatment as well as the underlying research methodology.

Authors Wulff and Gotzsche have produced an overall excellent book promoting the application of clinical research to everyday clinical practice and how one can learn from diagnostic, treatment and methodological mistakes as well as new advances. The book provides a beautiful and easy integration of statistical methods and medicine with everyday examples from clinical practice. The book provides the reader with a thoughtful rationale for basing clinical decisions on well conducted clinical research. I strongly recommend this excellent book to clinicians not actively engaged in research, medical students and health educators.

Moussa Sarr, MD, MPH
Health Studies Sector
WESTAT
Rockville, Maryland

PII S0197-2456(00)00107-0

INTRODUCTION

Posttraumatic stress disorder (PTSD) is a significant problem for a large number of veterans who receive treatment from the Department of Veterans Affairs (VA) health-care system. In an attempt to meet the need for PTSD treatment among veterans, the VA has funded over 140 specialized inpatient and outpatient PTSD programs across the country. The cost of these programs in 1998 was over $63.3 million, not including the cost of treatment received in general psychiatry programs [1].

The majority of VA PTSD patients are male Vietnam combat veterans. Their PTSD tends to be chronic, and many have significant psychiatric and psychosocial impairments. For example, 43% of veterans who sought treatment from the VA's specialized outpatient programs in 1998 had comorbid substance abuse, 68% were not working, and 55% were receiving compensation for problems related to PTSD [1]. Other studies have observed notable cognitive and physical impairments in veterans who seek VA care [2,3].

In 1995, the VA Cooperative Studies Program (CSP) approved CSP 420, a ten-site randomized clinical trial of group psychotherapy for treating PTSD in male Vietnam veterans. With a sample size of 360, the study was the largest randomized trial ever conducted of psychotherapy for PTSD and is one of the largest studies of group therapy for any disorder. Data collection was completed in June 2000. This article describes the design and methods of the study. The results will be reported elsewhere.

CHOICE OF THE EXPERIMENTAL TREATMENT

Planning for the study began with an examination of findings about the effectiveness of various treatments for PTSD. At the time, there had been relatively few randomized clinical trials of either pharmacotherapy or psychotherapy for PTSD [4]. However, the existing data indicated that psychotherapy was somewhat better than pharmacotherapy and that psychotherapy treatments involving some form of the cognitive-behavioral technique of exposure were particularly effective. It appeared that the most promising treatment to test in a large study would have to involve exposure therapy in some format, a conclusion supported by more recent meta-analytic results [5].

Another part of the planning process was to examine program evaluation data on the nature of treatment being delivered in the VA. These data showed that veterans received an average of 22.4 treatment sessions over 5.7 months, including 10.6 individual sessions and 11.8 group therapy sessions [6]. Most (65%) also received psychotropic medications. Despite this substantial amount of treatment, improvements in symptoms and functioning were modest and had questionable clinical significance [7]. Exposure-based treatment was delivered to less than 20% of the patients and was the primary treatment in only 1% of cases [6]. Dealing with war traumas constituted a major focus of treatment for only 11.4% and abreaction or deconditioning for less than 5% each.

Thus, wider use of exposure therapy for treating combat-related PTSD in VA patients seemed to be indicated. However, the likelihood of complications due to exposure is elevated in chronic, combat-related PTSD relative to other conditions [8]. VA patients have a number of risk factors that dissuade expert

clinicians from using exposure therapy [9], including excessive reactivity to traumatic memories, comorbid conditions such as alcoholism or personality disorder, and poor physical (especially cardiovascular) health.

The planning committee for CSP 420 chose to test a form of exposure therapy that was designed to maximize successful delivery to patients who might not otherwise be able to tolerate or comply with individual exposure therapy. The specific treatment, trauma focus group therapy (TFGT) [10], embeds exposure in a group context that includes psychoeducation, cognitive restructuring, a developmental perspective, relapse prevention, and coping skills training. The group approach provides patients with the dual opportunities for repeated exposure to their own war-related traumatic events as well as vicarious exposure to the traumatic events of other group members. Furthermore, groups promote healing by normalizing symptoms, increasing therapeutic opportunities, increasing generalizability of skill acquisition, and improving self-esteem by allowing each member to serve as helper as well as being helped [11]. In this environment there is a perception of safety that aims to increase the capacity of each patient to tolerate exposure.

CHOICE OF THE CONTROL TREATMENT

The standard control in a drug study is a placebo that appears identical to the drug under investigation. In a psychotherapy study, there is no single control condition that serves the same purpose. Thus, the question is not "what is the appropriate control group for evaluating a new psychotherapy?" but rather, "what inferences can be drawn from an experiment that employs a particular control group?"

CSP 420 used a nonspecific comparison design [12] to evaluate the effectiveness of TFGT. The control group in such a design is constructed to control for the "nonspecific" effects of psychotherapy—factors that all therapies have in common, such as therapist contact, instillation of hope, and expectation of improvement. A nonspecific comparison design therefore provides information about the mechanism behind a given therapy's effectiveness. If an experimental therapy is shown to be effective relative to a placebo therapy or care as usual, there is greater certainty that the effectiveness is due to specific aspects of the experimental therapy and not merely to nonspecific therapeutic factors.

The comparison treatment in CSP 420, present centered group therapy (PCGT), was designed to be a credible and clinically acceptable treatment condition. It is characterized by "nonspecific" and supportive kinds of interventions to control for the nonspecific benefits of the group experience. It does not include the exposure, restructuring, or other components that form the basis of TFGT. PCGT draws heavily from Yalom's [11] model of group therapy, which uses a "here-and-now" focus and emphasizes the process of interpersonal learning, group cohesion, and group support.

METHODS

Participants

The participants were 360 male veterans with PTSD due to service in the Vietnam theater. They had to consent to be randomized into treatment and agree

to terminate other psychotherapeutic treatment other than 12-step programs or pharmacotherapy during the trial. Those who were taking psychoactive medications had to be on a stable treatment regimen (i.e., no changes of drugs or dose) for a minimum of 2 months prior to entering the trial. However, once participants were enrolled, medication changes were permitted if clinically justified. In such cases, the change was documented for use in data analysis.

Psychotherapy research has been criticized for overly restricting the type of participants who can enter into clinical trials [13], thereby limiting the generalizability of findings. Inclusion and exclusion criteria for CSP 420 were designed to allow as broadly representative a sample as possible into the study. In particular, personality disorders and current substance abuse were not excluded, because these conditions are frequently comorbid with PTSD in VA patients [7].

Exclusion criteria included current alcohol or drug dependence, unwillingness to refrain from substance abuse at treatment or work, current or lifetime DSM-IV psychotic disorder, current major depression with psychotic features, current or lifetime mania or bipolar disorder, significant cognitive impairment, or a cardiovascular disorder that was judged by a cardiologist to prevent participation in the exposure component of the TFGT.

Assessment

The decision of what to measure in a clinical trial can be one of the most difficult to make, particularly for a study that is a substantial investment of resources. With a large sample, the inclusion of even one questionnaire measured at pretreatment, posttreatment, and follow-up can create many total hours of work for study personnel and participants alike. Given the size of CSP 420, measures were chosen based on their quality and the extent to which they were judged to be essential for broadly representing likely outcomes. Because TFGT so specifically targets PTSD symptoms, relatively little attention was paid to assessing other types of symptoms so that outcomes like functional status and quality of life could be assessed without creating excessive participant burden. An important consideration in the selection of a specific measure was the comparability it would facilitate to other relevant samples, such as individuals who seek treatment from VA's specialized outpatient PTSD programs [1].

Four domains were assessed: (1) psychological factors, including PTSD, substance abuse, and general distress; (2) psychosocial function, including an individual's work status, marital functioning, social/interpersonal functioning, legal status, and quality of life; (3) self-reported physical health status; and (4) utilization of physical and mental health services.

The primary outcome measure was PTSD severity as measured by the Clinician Administered PTSD Scale (CAPS) [14], a clinician-administered interview that reflects the diagnostic criteria in DSM-IV [15] and has excellent reliability and validity [16]. Other measures included the PTSD Checklist [17]; Mississippi Scale for Combat-Related PTSD [18]; Combat Exposure Scale [19]; Structured Clinical Interview for DSM-IV (SCID), patient version [20], including both Axis I and Axis II assessments and the Global Assessment of Functioning scale; 12-item version of the General Health Questionnaire [21]; Addiction Severity Index [22]; SF-36, the Medical Outcomes Study's Health Status Questionnaire [23]; Quality of Life Inventory [24]; and measures of physical and mental health

Table 1 Schedule of Assessments by Domain

	Months on Study						
Domain	Screening	0	1–5	7	12	18	24
PTSD/psychological							
CAPS	X	X		X	X	X	X
ASI (subscales)		X		X	X	X	X
GHQ		X	X	X	X		
PCL		X	X	X	X	X	X
SCID	X						
Mississippi Scale		X					
Combat Exposure Scale		X					
Psychosocial function							
ASI (subscales)		X		X	X	X	X
QOLI		X		X	X		
SF-36 (subscales)		X		X	X		
Global assessment	X						
Physical health							
SF-36 (subscales)		X		X	X		
Utilization		X		X	X	X	X

CAPS = Clinican-Administered PTSD Scale; ASI = Addition Severity Index; PCL = PTSD Checklist; GHQ = General Health Questionnaire; SCID = Structured Clinical Interview for DSM-IV; QOLI = Quality of Life Inventory.

service utilization used in the evaluation of VA's PTSD outpatient programs [1,7].

Table 1 lists the assessment schedule. Interviewers who were blind to participants' treatment condition performed assessments at study entry, the end of treatment (7 months), and the end of booster sessions (12 months). In addition, two-thirds of participants were assessed at 18 months and one-third were assessed at 24 months following study entry. Participants completed questionnaires monthly during the 7 months of active treatment. The decision to vary systematically the number of assessments received by subgroups of participants was made to optimize the competing aims of obtaining long-term follow-up data and minimizing the cost of data collection. Such longitudinal data, missing "by design," are amenable to analytic methods that are appropriate for unbalanced data.

All assessments were audiotaped. An independent clinician checked 8.33% of CAPS tapes ($n = 120$) and 25% of SCID tapes ($n = 90$) for reliability. Feedback was given to maintain consistency within and across interviewers over the course of the study.

Procedure

At each of the ten study sites, participants were referred to a masters- or doctoral-level interviewer by clinical staff. These staff were informed about the eligibility criteria and were encouraged in individual contacts and group settings to refer potentially eligible participants—a process that was repeated throughout the study.

Screening information was obtained in three phases, structured to minimize both participant burden and unnecessary cost to the study due to extensive

assessment of ineligible participants. In the first phase, the referral source was consulted to establish provisional psychiatric diagnoses, and patient records were searched to confirm that the veteran had served in the Vietnam War theater. The second phase consisted of an interview that contained demographic questions intended to facilitate comparisons of potential participants who were ruled in versus ruled out of the study, questions about cardiovascular health, a brief assessment of cognitive function [25], and questions about willingness to adhere to study conditions. Prior to being accepted into the last phase of screening, potential participants who reported any indication of cardiovascular problems were referred to a cardiologist to determine whether the problems would make it dangerous for the participant to tolerate the physiological arousal that can occur during the exposure component of TFGT.

Of the 563 potential participants who were contacted and invited to enter the second phase of screening, 16.5% (n = 93) refused or were unable to participate: 6.2% would not give consent, 4.8% felt they were unable to participate (e.g., due to study requirements, scheduling conflicts), 0.7% did not like TFGT, and 4.8% gave no reason.

During the second phase of screening, the interviewer reviewed an informed consent form with potential participants to explain the study in more detail. TFGT was described in neutral language to minimize the demand for improvement in TFGT:

> You are being asked to participate in a research treatment program testing two types of group therapy for Post Traumatic Stress Disorder (or PTSD). People who have PTSD are often troubled by memories of past traumatic events, nervousness, depression, and feeling distant from others. One kind of therapy, Trauma Focus Group Therapy (or TFGT), involves focusing on memories of your Vietnam war-zone experiences and helping you develop specific ways to deal with them. The other kind of therapy, Present Centered Group Therapy (PCGT), involves talking about day-to-day problem behaviors and feelings that interfere with your present life without going over your Vietnam war-zone experiences. . . . Since it is not known which treatment will be most helpful in reducing symptoms of PTSD, you will be assigned at random (like a flip of a coin) to one of the two treatment groups, TFGT or PCGT.

However, participants were not asked to sign the consent form until they returned for diagnostic interviewing in the third phase of screening. In this last phase, potential participants who met the criteria assessed in the first two phases underwent a structured psychiatric interview to assess inclusion and exclusion diagnoses.

Because both TFGT and PCGT are group rather than individual therapies, it was necessary to accrue 12 participants for a cohort at each study site before treatment could begin. The 12 participants were then randomized to either TFGT or PCGT in groups of six each, with randomization stratified by CAPS severity. There were three cohorts per site. Each successive cohort began active treatment 1 month after the preceding cohort had begun booster sessions.

A group-based design such as this delays treatment for enrolled participants who must wait while a cohort is being assembled. To minimize the amount of time participants had to wait for a group to begin, recruitment and screening phases were very high intensity in terms of workload. Staffing sites with two

full-time research positions, one concentrating on recruitment and scheduling, the other concentrating on recruitment and assessment, enabled the sites to meet this workload. Prior to enrollment for a given cohort, study staff worked with referral sources to develop a list of potential eligible participants. Formal screening of referral sources in the first phase of our screening process provided a quick and efficient way to rule out ineligible participants so that more time could be allotted to in-person assessments of potential participants. Case management was begun as soon as a participant was enrolled to minimize dropout during this holding period and to provide interim clinical care (because participants were required to give up most forms of therapy).

We believe these retention strategies were successful: only 3.7% of participants (n = 21) dropped out prior to randomization after meeting eligibility criteria. The most common reasons for dropout during this period were schedule conflict (n = 5), medication change (n = 4), and no reason (n = 4).

In a clinical trial that requires participants to give up treatments other than those delivered in the trial, case management can serve as a point of individual contact for each participant so that adequate monitoring of clinical status is ensured and assistance with additional services (e.g., medical, legal, financial) can be provided. By addressing individual participant needs as they arise, case management can allow the work of group treatments to remain consistent and focused. Case management in CSP 420 was delivered according to a manualized protocol. There were two or three case managers per site, and they met with participants weekly before treatment began, every other week during the first month of treatment, and then monthly throughout each participant's total time in the study, i.e., 12, 18, or 24 months. Additional visits were allowed when clinically indicated, as were additional inpatient or outpatient treatment for clinical emergencies. The provision of additional treatment was recorded for use in data analysis.

Treatment Delivery

TFGT and PCGT treatment were delivered in weekly sessions for 30 weeks according a manualized protocol for each treatment. All sessions lasted 1.5 hours, except for exposure sessions, which lasted 2 hours. Monthly 1.5-hour group booster sessions were delivered for the 5 months following active treatment; 15-minute booster phone cells also were delivered monthly during this period for the TFGT condition.

Two therapists led each group. To participate in the study, therapists had to be masters- or doctoral-level clinicians with prior experience in treating PTSD in a group format. They were not required to have formal training in exposure techniques or even in cognitive-behavioral therapy. The use of nonexperts will enhance the generalizability of findings to possible real-world conditions: TFGT being delivered by therapists new to the technique and not just TFGT as delivered by experts.

The problem of therapist effects often comes up in psychotherapy research. The first design decision is whether to find experts in each therapy under study and assign each of them to deliver that therapy alone ("specialize"). If the therapies under study are to be delivered in clinical practice only by experts, then this is an appealing choice, because it matches the study design to the

target clinical practice. However, if experts in one of the study treatments are more skilled therapists than experts in the other, then such a design confounds skill with treatment. We worried that this was likely to be true in CSP 420, because one of the therapies, TFGT, was a "special" experimental therapy, and expertise in TFGT might be expected to be associated with experience, training, and other markers of skill. Furthermore, we thought it likely that TFGT would be delivered by a broad range of therapists in clinical practice if it proved successful in CSP 420. Having decided not to specialize in CSP 420, we could have assigned each therapist to deliver both treatments in a "counterbalanced" or crossover design. The counterbalanced design can remove therapist effects from the comparison of treatments, insofar as therapist preferences or skills do not cause treatments to be delivered with different levels of enthusiasm or competence. However, we rejected this approach because of a concern that having therapists deliver both treatments would make it hard for them to keep the treatments distinct in application. Instead, to ensure that therapist effects were balanced across treatments, we randomized therapists to the single treatment each was to deliver.

All sessions were videotaped. Telephone supervision based on the tapes was provided weekly by a senior therapist chosen to monitor the therapy in each condition. Global ratings of protocol adherence and therapist competence were made for all tapes in each condition by the senior therapist for that condition. In addition, three sessions from each of the 60 groups run during the study ($n = 180$, 17%) were monitored for adherence by two senior clinicians who were independent of the treatment delivery. The independent raters also made global ratings of adherence and competence and rated specific elements to provide a manipulation check, e.g., to ensure that trauma focus or cognitive restructuring did not occur in PCGT sessions.

Statistical Issues

The outcome of treatment was measured at several points in time for each individual. The pivotal comparison was immediately following the end of treatment (7 months) because this was the time when improvements were expected to be most evident. In addition, the percentage of missing data was likely to be lowest. Analyses of subsequent outcomes will concentrate on the question of the durability of effects, assuming they are present at the end of acute treatment. Thus, the main analysis is a cross-sectional comparison of results at a fixed point in time rather than a longitudinal profile comparison.

For sample size estimation, an effect size of $d = 0.5$ was judged to be the minimum effect that would be clinically meaningful. A difference of 0.5 SDs represents a decrease of approximately ten points on the CAPS in treatment-seeking Vietnam veterans with PTSD, for whom the SD is roughly 20 [7]. Ten points can represent meaningful improvement in the life of someone with chronic PTSD. For example, if a person had ten of the 17 PTSD symptoms at maximum intensity and frequency, it could mean having all symptoms decline: in frequency, from daily to four or five times a week, or in intensity, from complete incapacity to some functional ability. For a single symptom like nightmares, this could mean going from traumatic nightmares that prevent

one from returning to sleep every night to getting two or three nights of sleep per week uninterrupted by nightmares.

It is possible that dropouts from treatment received care outside the study that resembled the treatment to which they were not assigned, e.g., a TFGT dropout may have received treatment that focused on his current interpersonal problems. Under the intent-to-treat principle, these noncompliant participants will be counted in their original treatment groups. This tends to reduce the apparent effect size and thus lowers the power to detect a true effect size of 0.5.

Analysis by intention-to-treat implies complete follow-up of outcomes in participants, and this in turn requires their cooperation with measurement. A few participants did refuse to be assessed at the critical 7-month visit. While we made strenuous efforts to obtain full follow-up in all participants regardless of adherence to study treatment, research ethics dictated that we honor a participant's refusal to be interviewed. We will use statistical methods for missing data, such as imputation based on "missing at random" assumptions [26], to assess the potential impact of the expected small number of missing observations. We emphasize that these statistical methods cannot compensate for substantially incomplete data collection, but can provide some reassurance that the conclusions of the study would not be changed by observing them. Because missing data methods rely on untestable hypotheses (usually about the missing data mechanism), in the conservative world of clinical trials they are appropriately used in the form of sensitivity analyses.

Clustering

As described above, each site accrued 12 participants in each cohort before participants were stratified by CAPS severity and randomized to a treatment condition. The randomizations were done for each participant using permuted blocks of four in three blocks of CAPS severity scores to ensure the balance of treatment groups by CAPS score. The six participants assigned to each treatment arm formed a "cluster" for group therapy.

A correct analysis of the data must take the cluster effect into account. Even though patients are randomized individually, the analysis still must take the cluster effect into account because of the correlation of outcomes within groups. This correlation is ρ, the intraclass correlation coefficient. There is little published evidence to suggest what the typical, or expected, intraclass correlation among group members might be, but a recent study of group psychotherapy for PTSD found intraclass correlations for PTSD outcomes ranging from 0.096 to 0.131 [27]. Assuming that this correlation is 0.0 can have negative consequences for statistical power as discussed below.

There are potentially two levels of clustering in this design. The clustering due to the group treatments is more likely to cause difficulty in the analysis, because each therapy group receives one treatment, while both treatments occur at each site. Therefore, we adjusted the sample size to account for treatment group clustering and propose to handle the site effects in the analysis.

Inflation of Variance

Suppose $n_{\text{group}} = mk$ participants are assigned to k groups, m per group, then a direct calculation yields:

$$\mathrm{Var}(\overline{Y}_{\mathrm{Group}}) = \frac{\mathrm{Var}(Y_{\mathrm{Group}})}{mk}[1 + (m - 1)\rho] = \frac{\sigma_G^2}{mk}f \tag{1}$$

where ρ is the intraclass correlation coefficient:

$$\rho = \frac{\sigma_b^2}{\mathrm{Var}(Y_{\mathrm{Group}})} \tag{2}$$

the ratio of the "between-groups" variance to the total variance, which is the sum of the between- and within-group variance:

$$\mathrm{Var}(Y_{\mathrm{Group}}) = \sigma_b^2 + \sigma_w^2 \tag{3}$$

The first term on the right-hand side of Eq. (1) is just the variance of the mean of $n_{\mathrm{group}} = mk$ independent measures, so the second term $f = 1 + (m - 1)\rho$ can be interpreted as the "inflation factor" for the variance [28]. It is called a "design effect" in survey science, where Eq. (1) is a familiar equation. The adjustment for design effects is well established for many situations that generalize the mean, such as estimation of fixed effects (e.g., treatment differences, regression coefficients, etc.) in linear models, and has been extended to generalized linear models recently.

Sample Size Inflation

The design of this study differs from most group intervention studies where both the unit of randomization and intervention are a cluster or group of participants. However, sample size calculation for this individually randomized study is the same as cluster randomization with a cluster size of $m = 6$. Although there were complete sample size formulas available for cluster randomization, a much simpler method can be used if the intraclass correlation coefficient of the outcome data can be reasonably assumed [28]. As a consequence of variance inflation, the sample size needs to be increased by the same inflation factor f to achieve the variance reduction that one would have anticipated. Some features to note are: (1) if the group size is just one patient, m is 1 and there is no inflation ($f = 1$); and (2) if the intraclass correlation is 1.0, then $f = 1 + (m - 1) = m$, indicating that the within-group averaging is no help at all and one needs to act as if the group is an individual. Table 2 presents examples of sample sizes that were calculated by assuming that the cost per cluster is just the cluster sizes times the individual cost. As can be seen from Table 2, even small intraclass correlations can have substantial effects on the needed sample size. If $m = 6$, an intraclass correlation of only 0.10 inflates the necessary sample size by 50%. Failing to include additional participants to compensate for this inflation could have substantial effects on type II error. Table 2 also illustrates that the effect of within-group correlation on sample size projections is increased at larger group sizes.

Researchers who employ group interventions need to consider the possible inflation of sample size requirements due to the correlation of outcomes within therapy groups. If the intraclass correlation is high, adding members to groups does not help with variance as much as adding groups. When the intraclass correlation is greater than 0.30 and the number of members per group is greater than ten, the variance is insensitive to the addition of new members to groups,

Table 2 Sample Size as a Function of Intraclass Correlation and Group Size

Intraclass Correlation ρ	Number of Patients per Group (m)	Inflation Factor $f = 1 + (m - 1)\rho$	Sample Size per Treatment Arm
0.00	6	1.00	64
0.05	6	1.25	80
0.10	6	1.50	96
0.20	6	2.00	128
0.00	11	1.00	64
0.05	11	1.50	96
0.10	11	2.00	128
0.20	11	3.00	192

The calculations listed above assume that desired power is 0.80 to find an effect size $d = 0.50$ at alpha = 0.05 (two-tailed).

and thus, so is the power of the test [28]. The sample size inflation is based on univariate analysis using comparisons at specific time points at each site. Multivariate analysis to adjust for clustering and other confounding variables will be used for the longitudinal data and are considered as secondary analyses.

DISCUSSION

Seligman [13] generated a great deal of controversy when he argued for the merits of "effectiveness" studies without rigorous scientific control on the grounds that controlled "efficacy" studies had extremely limited generalizability to real-world settings. He proposed that the ideal psychotherapy study "would combine several of the best features of efficacy studies with the realism of the survey [effectiveness] method." We agree wholeheartedly and have attempted to do so with various aspects of the design of CSP 420: relatively broad inclusion criteria, the use of nonexpert therapists, not dropping participants who need additional treatment, and an assessment battery that comprehensively measures treatment outcome.

Designing and implementing CSP 420 raised a number of general issues concerning large-scale multisite psychotherapy research. In addition to the logistic and methodological challenges of any multisite study, psychotherapy research presents complex issues that are not encountered in medication trials. Whereas industrial pharmaceutical quality control ensures that all dispensed pharmacological agents are comparable in a medication trial, no such assurance can be built into a psychotherapy trial. In CSP 420, we attempted to optimize psychotherapy quality control through a number of features: (1) designing detailed treatment manuals for both active and control interventions; (2) training all therapists to an acceptable level of competence through formal training and, in the active condition, through clinical supervision of a pilot group; (3) ongoing monitoring of all treatment by videotaping each session; (4) providing weekly supervision of all therapists by master therapists; and (5) using master raters to assess adherence and fidelity with a detailed monitoring scale.

Some of these issues, such as maintaining protocol adherence across sites, are similar to those encountered in the National Institute of Mental Health Treatment of Depression Collaborative Research Program (TDCRP) [29]. However, the size of CSP 420 made these issues particularly challenging. Whereas TDCRP randomized 240 participants to individual therapy at three sites, CSP 420 randomized 360 patients to 60 groups across ten sites. To our knowledge, CSP 420 is one of the largest psychotherapy studies that the VA has ever funded. The study would not be possible without the help of the VA Cooperative Studies Program. The program is designed to support multisite trials and capitalizes on the multisite nature of the VA itself. A large and complicated study like CSP 420 could not be planned or run in any cost-effective manner without the support of the research infrastructure provided by the program.

Technological advances, and the expansion of internet communication, also greatly facilitated study management. The Cooperative Studies Program relies on fax-based transmission of data, making centralized data collection much faster and more efficient than it might be through other means. The use of e-mail allowed rapid multisite communication and dissemination of study materials, such as assessment instruments and changes to the operations manual. Although we were not able to use videoconferencing for meetings and supervision, investigators should consider this for future studies. Recording of assessment and treatment sessions for fidelity and process monitoring also might be facilitated by reliance on new technologies. CSP 420 generated over 2000 videotapes and 10,000 audiotapes, all of which were duplicated and mailed to central study locations for reliability monitoring and assessment of manual adherence. With computerized recording, we could have reduced the costs associated with duplicating, mailing, and storing such a large number of tapes.

Another aspect of CSP 420 that provided challenges in addition to those encountered in multisite treatment research was the decision to study group rather than individual therapy. One challenge was recruitment, which required that participants accepted into the study wait until we had identified enough participants to fill two therapy groups. As indicated above, recruitment and screening created significant pressure for study staff. The waiting period also was potentially difficult for participants because they were required to terminate most forms of ongoing therapy. We initiated case management at this time to respond to their clinical needs, with the hope of preventing dropout before participants entered active treatment. Another challenge was determining an optimal number of therapy sessions. On the one hand, it was important that the total number of sessions be realistic and comparable in both conditions. On the other, it was necessary to ensure that participants assigned to TFGT had a sufficient amount of exposure. We addressed both needs by providing two in-group exposure sessions and audiotaping each veteran's exposure sessions, to be listened to at least eight times as homework. There were minor logistic challenges as well, such as finding a time when all group members could attend a session when individuals would not know their group assignment until after randomization.

However, the most significant challenges came from the effects of group clustering on sample size estimation and data analysis. The delivery of a group intervention required substantially more participants than would have been

required for the evaluation of an individual treatment. Data analysis also needs to reflect the group clustering, although a recent review indicates that almost 90% of studies of group therapy have failed to do so [30].

The TDCRP [29] reflected a groundbreaking effort that has substantially shaped the design and implementation of multisite psychotherapy research, including CSP 420. It is our hope that sharing the details of CSP 420 builds upon this remarkable contribution by providing information relevant to multisite trials of group interventions. In addition to these methodological implications, the substantive findings of CSP 420 will provide information relevant to the treatment of chronic PTSD, as well as a unique, group-based model of exposure therapy [10]. Many VA patients exhibit a high level of symptoms and functional difficulties even after treatment [7]. If TFGT is shown to be effective, it may be useful for treating PTSD in other chronic populations. Also, by demonstrating the efficacy of exposure therapy when delivered in a group context, CSP 420 will expand the available tools for addressing the long-term sequelae of traumatic exposure.

This research was supported by a grant to M.J. Friedman and P.P. Schnurr from the VA Cooperative Studies Program. The current project manager is Nancy Bernardy, PhD. The former project manager is Beth H. Stamm, PhD. Therapy supervision was provided by Shirley M. Glynn, PhD, and Melissa Wattenberg, PhD, and case management supervision was provided by Philip Massad, PhD. Robert Rosenheck, MD, and Raymond M. Scurfield, DSW, participated in the study planning, and Josef I. Ruzek, PhD, and Sherry J. Riney, MSW, contributed to the development of treatment manuals. Numerous staff from the VA Cooperative Studies Program have contributed to aspects of the study planning and implementation, particularly Andrea Millman, MA, and Kathy Small.

REFERENCES

1. Fontana A, Rosenheck R, Spencer H, Gray S, DiLella D. *The Long Journey Home, VII: Treatment of PTSD in the Department of Veterans Affairs: Fiscal Year 1998 Service Delivery and Performance.* West Haven, CT: VA Northeast Program Evaluation Center and National Center for PTSD; 1999.
2. Beckham JC, Moore SD, Feldman ME, et al. Health status, somatization, and severity of posttraumatic stress disorder in Vietnam combat veterans with posttraumatic stress disorder. *Am J Psychiatry* 1998;155:1565–1569.
3. Bremner JD, Scott TM, Delane RC, et al. Deficits in short-term memory in posttraumatic stress disorder. *Am J Psychiatry* 1993;150:1015–1019.
4. Solomon SD, Gerrity ET, Muff AM. Efficacy of treatments for posttraumatic stress disorder: An empirical review. *JAMA* 1992;268:633–638.
5. Van Etten ML, Taylor S. Comparative efficacy of treatments for post-traumatic stress disorder: A meta-analysis. *Clin Psychol Psychother* 1998;5:126–144.
6. Fontana A, Rosenheck R, Spencer H. *The Long Journey Home: The First Progress Report on the Department of Veterans Affairs PTSD Clinical Teams Program.* West Haven, CT: VA Northeast Program Evaluation Center and National Center for PTSD; 1990.
7. Fontana A, Rosenheck R. Effectiveness and cost of the inpatient treatment of posttraumatic stress disorder: Comparison of three models of treatment. *Am J Psychiatry* 1997;154:758–765.
8. Foy DW, Kagan B, McDermott C, et al. Practical parameters in the use of flooding for treating chronic PTSD. *Clin Psychol Psychother* 1996;3:169–175.
9. Litz BT, Blake DD, Gerardi RG, Keane TM. Decision making guidelines for the use of Direct Therapeutic Exposure in the treatment of Post-Traumatic Stress Disorder. *Behav Therapist* 1990;13:91–93.

10. Foy DW, Glynn SM, Ruzek JI, Riney SJ, Gusman FD. Trauma focus group therapy for combat-related PTSD. *In Session: Psychotherapy in Practice* 1997;3:59–73.
11. Yalom ID. *The Theory and Practice of Group Psychotherapy*. 3rd ed. New York: Basic Books; 1985.
12. Borkovec TD. Between-group therapy outcome research: Design and methodology. *NIDA Research Monographs* 1993;137:249–289.
13. Seligman ME. The effectiveness of psychotherapy. The Consumer Reports study. *Am Psychol* 1995;50:965–974.
14. Blake DD, Weathers FW, Nagy LM, et al. The development of a Clinician-Administered PTSD Scale. *J Trauma Stress* 1995;8:75–90.
15. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 4th ed. Washington, DC: American Psychiatric Association; 1994.
16. Weathers FW, Ruscio AM, Keane TM. Psychometric properties of nine scoring rules for the Clinician-Administered Posttraumatic Stress Disorder Scale. *Psychol Assess* 1999;11:124–133.
17. Weathers FW, Litz BT, Herman DS, Huska JA, Keane TM. PTSD Checklist: Reliability, validity, and diagnostic utility. In: Proceedings of the 9th Annual Meeting of the International Society for Traumatic Stress Studies (ISTSS). Chicago, IL: ISTSS; 1993:8.
18. Keane TM, Caddell JM, Taylor KL. Mississippi Scale for Combat-Related Posttraumatic Stress Disorder: Three studies in reliability and validity. *J Consult Clin Psychol* 1988;56:85–90.
19. Keane TM, Fairbank JA, Caddell JM, et al. Clinical evaluation of a measure to assess combat exposure. *Psychol Assess* 1989;1:53–55.
20. Spitzer RL, Williams JB, Gibbon M, First MB. Structured Clinical Interview for DSM-IV—Patient Version (SCID-P, Version 2.0). Washington, DC: American Psychiatric Press; 1995.
21. McFarlane AC. Long-term psychiatric morbidity after a natural disaster: Implications for disaster planners and emergency services. *Med J Aust* 1986;145:561–563.
22. McLellan A, Kushner H, Metzger D, et al. The fifth edition of the Addiction Severity Index. *J Subst Abuse Treat* 1992;9:199–213.
23. Ware JE, Sherbourne CD. The MOS 36-Item Short-Form Health Survey (SF-36): I. Conceptual framework and item selection. *Med Care* 1992;30:473–483.
24. Frisch MB, Cornell J, Villanueva M, Retzlaff PJ. Clinical validation of the Quality of Life Inventory: A measure of life satisfaction for use in treatment planning and outcome assessment. *Psychol Assess* 1992;4:92–101.
25. Wechsler D. *Wechsler Memory Scale-Revised Manual*. San Antonio, TX: The Psychological Corporation; 1987.
26. Little RJA, Rubin DB. *Statistical Analysis with Missing Data*. New York, NY: Wiley; 1987.
27. Creamer M, Morris P, Biddle D, Elliott P. Treatment outcome in combat-related PTSD: A cause for cautious optimism. *J Trauma Stress* 1999;12:545–558.
28. Hsieh FY. Sample size formulae for intervention studies with the cluster as unit of randomization. *Stat Med* 1988;7:1195–1201.
29. Elkin I, Parloff MB, Hadley SW, Autry JH. NIMH Treatment of Depression Collaborative Research Program. Background and research plan. *Arch Gen Psychiatry* 1985;42: 305–316.
30. Burlingame GM, Kircher JC, Taylor S. In: Fuhriman A, Burlingame GM, eds. Methodological considerations in group psychotherapy research: Past, present, and future practices. *Handbook of group psychotherapy: An Empirical and Clinical Synthesis*. New York, NY: Wiley; 1994:41–80.

APPENDIX

Key Personnel for CSP 420: Group Treatment of PTSD

Coprincipal Investigators

Matthew J. Friedman, MD, PhD, National Center for PTSD, White River Junction VAM&ROC, and Department of Psychiatry and Pharmacology, Dartmouth Medical School

Paula P. Schnurr, PhD, National Center for PTSD, White River Junction VAM&ROC, and Departments of Psychiatry, Dartmouth Medical School

Executive Committee Members

David W. Foy, PhD, Graduate School of Education & Psychology, Pepperdine University

Fred D. Gusman, MSW, National Center for PTSD, VA Palo Alto Healthcare System

Terence M. Keane, PhD, National Center for PTSD, Boston VAMC, and Department of Psychiatry, Boston University School of Medicine

Charles R. Marmar, MD, San Francisco VAMC, and Department of Psychiatry, University of California at San Francisco

Bruce J. Rounsaville, MD, Department of Psychiatry, Yale University School of Medicine

M. Tracie Shea, PhD, Providence VAMC, and Department of Psychiatry, Brown University School of Medicine

Daniel S. Weiss, PhD, San Francisco VAMC, and Department of Psychiatry, University of California at San Francisco

VA Cooperative Studies Participants

Frank Y. Hsieh, PhD, VA Cooperative Studies Program, VA Palo Alto Healthcare System

Philip W. Lavori, PhD, VA Cooperative Studies Program, VA Palo Alto Healthcare System, and Department of Health Research and Policy, Stanford University