Advancing Research Standards for PTSD Interventions: Suggested Approaches for Designing and Evaluating Clinical Trials

A Meeting Summary

Convened by:
U.S. Department of Veterans Affairs Office of Research & Development in collaboration with:
National Institute of Mental Health and
U.S. Department of Defense

January 22-23, 2008
Rockville, Maryland

Advancing Research Standards for PTSD Interventions: Suggested Approaches for Designing and Evaluating Clinical Trials

A Meeting Summary

Sponsored by the U.S. Department of Veterans Affairs in collaboration with the National Institute of Mental Health and the Department of Defense, Office of the Secretary of Defense for Health Affairs

Objective: On January 22 and 23, 2008, the U.S. Department of Veterans Affairs (VA), the National Institute of Mental Health (NIMH), and offices of the Department of Defense (DoD) (Office of the Secretary of Defense for Health Affairs, Defense Center of Excellence, and Congressionally Mandated Research Program) convened a group of scientific experts and research administrators to develop approaches for overcoming challenges in conducting rigorous interventional research in persons with posttraumatic stress disorder (PTSD). **Method:** Experts were asked to consider issues raised by the Institute of Medicine (IOM) report titled *Treatment of PTSD: An Assessment of the Evidence* (2008) and to address specific questions related to clinical trial design and analysis; the selection, use, and interpretation of measures designed to assess PTSD; and trial implementation. At the 2-day meeting, facilitators led discussions of questions developed by organizers, and experts deliberated to develop suggested approaches. **Results:** Suggestions were developed for clinical trial design and analysis, PTSD measurement, and trial implementation. Specific suggested approaches included alignment of study design and appropriate endpoints with study objectives, adherence to study protocol, broad inclusion criteria for study participants, and appropriate selection of control groups. Additional discussion focused on approaches to including subjects with incomplete data in study results, the importance of applying prespecified strategies for addressing missing data and strategies for followup investigations, the use of self-report measures, and subject reimbursement. **Conclusions:** Many of the suggested approaches may be applicable to psychiatric interventional research globally. In the case of PTSD, these approaches could provide a foundation for researchers and scientific reviewers to use in designing and evaluating future interventional studies, thereby strengthening the evidence base for PTSD treatment approaches across multiple populations.

Introduction

Many reports on the prevalence of PTSD among veterans and civilian trauma survivors indicate that PTSD is a serious public health concern. Yet, a 2008 IOM evaluation suggested there is inconclusive evidence for the effectiveness of most pharmacologic and psychosocial interventions for PTSD. According to the IOM report, this may be due to formidable challenges around conducting PTSD intervention research. For example, comorbid mood, anxiety, and substance use disorders (SUDs) often complicate diagnosis and in some instances make it difficult to determine whether treatment directly impacts PTSD. Further, concurrent clinical treatment for comorbid physical and psychological conditions may interfere with the mechanisms of PTSD interventions. Differences of opinion about diagnosis and inclusion criteria

for PTSD trials present challenges as well, and little guidance exists for characterizing trauma exposures, choosing control conditions, and timing PTSD interventions. The IOM report therefore recommended that the VA and other Government agencies that fund clinical research ensure that PTSD interventional studies take steps toward addressing these and other problems that affect the quality of the research.

In response to the challenges identified by the IOM, specifically those related to methodology and design of trials, as well as the need to continue to advance the evidence base for effective PTSD treatments, the VA, NIMH, and DoD convened a group of experts to develop suggested approaches to the design and evaluation of PTSD clinical trials, to improve the quality of PTSD interventional research, and to enhance comparability between studies, thus providing a stronger evidence base for PTSD treatments. This paper summarizes the workgroup's suggested approaches, which address general aspects of designing rigorous, well-controlled trials along with specific issues related to PTSD interventional research.

Methods

Prior to the meeting, an organizing committee comprising VA, NIMH, and DoD scientists contacted experts in clinical trial design as well as clinicians in the field (see list at end of manuscript). These experts were asked to consider both IOM and other expert recommendations to enhance the evidence base for PTSD treatments and to address specific questions relating to PTSD interventional studies. Experts were assigned to one of three subgroups: (1) PTSD trial design and analysis, (2) PTSD measurement, and (3) PTSD trial implementation (a member of a subgroup may have expertise in multiple subgroups). Each subgroup held one to two telephone conferences prior to the meeting to identify research challenges under their topic area. The 2-day meeting that followed began with an overview of current standards for designing and conducting clinical trials, presentations on the complexities in PTSD trials, and examples of studies that surmounted the challenges associated with conducting PTSD interventional research. Subgroup facilitators then led discussions with the entire group of questions assigned to their subgroups, ensuring that suggested approaches were fully discussed and developed. Following the meeting, subgroups met again via conference call to review draft approaches. The results of these deliberations are the suggested approaches presented in this paper.

Results

This section identifies questions considered by the expert workgroup regarding (1) PTSD trial design and analysis, (2) PTSD measurement, and (3) PTSD trial implementation, with suggested approaches and accompanying rationale for addressing these issues in future research studies.

Suggestions for PTSD Trial Design and Analysis

Question 1: STUDY OBJECTIVES. What issues should be considered in specifying objectives, specific aims, hypotheses, and study design for PTSD clinical trials?

Suggested Approach: Objectives, specific aims, hypotheses, and study design should be clearly stated and justified in terms of the existing literature. Objectives and aims should be feasible, and

hypotheses should be stated in a testable format that is aligned with the data analysis plan. Further, primary, secondary, and exploratory hypotheses should be clearly distinguished from one another. The study design should be appropriate in terms of the state of knowledge about a topic as well as the intended goals for applying the study's findings.

Rationale: The Consolidated Standards of Reporting Trials, or CONSORT, states, "Objectives are the questions that the trial was designed to answer. They often relate to the efficacy of a particular therapeutic or preventive intervention. Hypotheses are pre-specified questions being tested to help meet the objectives. Hypotheses are more specific than objectives and are amenable to explicit statistical evaluation. In practice, objectives and hypotheses are not always easily differentiated..." (Altman et al. 2001, 134:669).

The appropriateness of pilot studies, single-site clinical trials, and/or multisite randomized controlled clinical trials depends on the research objective. Pilot studies are useful to establish feasibility, to refine procedures, and to implement proven treatments in settings or populations in which a treatment has not been used previously. Single-site trials typically would be used earlier in the research process to provide proof of concept or preliminary evidence of efficacy; e.g., estimates of safety, tolerability, and effect size. Multisite trials typically would be used to assess generalizability and to obtain practical information about the widespread application of the treatment. Multisite trials also may be needed for designs that require a high number of subjects; e.g., dismantling/additive designs, designed tests of mediators and moderators, and tests designed to detect small effects such as those seen in comparisons of two treatments that are each hypothesized to be active.

Question 2: INCLUSION/EXCLUSION CRITERIA AND COMORBIDITIES. What are key considerations for inclusion/exclusion criteria to ensure representativeness of the subject population and generalizability of the findings?

Suggested Approach: A study's sampling frame depends in part on the question being asked. Pilot studies looking for initial evidence of a "signal" may be more restrictive in terms of subject selection criteria, while larger effectiveness trials should be representative of the clinical population of interest and allow for generalizable conclusions. Broad inclusion criteria should be used in effectiveness studies to adequately represent the clinical populations at risk, including subjects with psychiatric and general medical comorbidities (e.g., depression, anxiety disorders, SUDs, personality disorders, traumatic brain injury [TBI]) and those with subthreshold PTSD with high symptom severity). Stratification variables should be chosen based on an established relationship with outcome and should rarely exceed one or two in number.

Rationale: More inclusive subject selection criteria facilitate recruitment and produce study samples that are more representative of community populations, resulting in wider generalizability. In traditional efficacy studies, subjects are eligible for treatment only if they meet restrictive inclusion/exclusion criteria that often eliminate subjects with comorbid mental disorders. However, epidemiological studies highlight that comorbidity is the rule rather than the exception, particularly with PTSD (Kessler et al. 2005). Rates of comorbidity ranging from 62 to 92 percent have been reported in population-based surveys of PTSD (Davidson et al. 1991).

First-generation investigational studies of a treatment should carefully measure comorbidity and check for randomization failure. If randomization failure is found, these studies should potentially include comorbidity as a covariate. Second-generation studies should continue in the same fashion unless and until evidence mounts that there is a confounding comorbidity that justifies stratification during the design phase. An excessive number of stratification variables can result in a sparse number of subjects in some strata, which can compromise both the randomization process and statistical power.

In certain situations, subjects may be excluded for any number of reasons, including safety of clinical care, significant cognitive impairment that suggests inability to give consent, presence of a psychotic or bipolar disorder not in remission, or imminent risk for suicide or violence. Subject characteristics should be explicitly described in PTSD treatment trials, particularly if used to determine study inclusion/exclusion so that representativeness of the study sample to the relevant patient population can be judged. These defined characteristics should include psychiatric and medical comorbidities, age, trauma type, cohort effects (e.g., for veterans, war zone, military era, etc.), race, gender, socioeconomic status, psychiatric medication status, co-occurring mental health treatment, and changes in treatments during the course of the study.

Question 3: COMPARISON GROUP. What are the characteristics of an appropriate comparison group to use in a randomized controlled clinical trial of psychotherapy for PTSD?

Suggested Approach: The comparison condition for a randomized controlled trial (RCT) should be selected based on the scientific question that the clinical trial is designed to answer. Other considerations when selecting the comparison condition include (1) practical and ethical issues, (2) the tradeoff between maximizing internal versus external validity, (3) the status of existing scientific knowledge about the experimental treatment, and (4) the phase in the overall progression of research into the treatment being studied. In effectiveness trials, comparing the new treatment to a "clinically relevant alternative" intervention is essential for the new treatment to be adopted by the field, particularly when an existing, less-expensive standard of practice is available (Tunis, Stryer, and Clancy 2003; Glasgow et al. 2005). A credible comparison group must account for the passage of time, increased attention, and the expectation of receiving an effective intervention. A waitlist comparison group should be used only in the initial stages of treatment development and trials.

Rationale: Some comparison conditions represent true controls for confounding factors that threaten internal validity (e.g., maturation) or construct validity (e.g., nonspecific therapeutic factors), while other comparison conditions represent other active forms of treatment in a superiority trial. Practical and ethical issues as well as the tradeoff between maximizing internal versus external validity need to be taken into consideration in choosing a comparison condition. Phases in the progression of clinical trials research that dictate appropriate choice of comparison conditions have been described extensively (Greenwald and Cullen 1985; Rounsaville, Carroll, and Onken 2001; Schnurr 2007; Mercer et al. 2007) and generally include pilot studies, efficacy trials, effectiveness trials, and translational research.

The comparison group used in a particular RCT may be chosen from a range of possibilities depending on the aforementioned stage of research addressed by the study question. Although

waitlist designs control for most threats to internal validity (Borkovec 1993; Schnurr 2007), they do not control for threats to construct validity. Thus, they maybe appropriate in the early stages of treatment development only. Once a treatment has been shown to be effective in a waitlist design, subsequent research should employ comparison groups that control at least for attention, therapist contact, and other nonspecific factors. After the efficacy of a treatment has been established, it is important to determine its effectiveness in "real world" clinical settings across a broad array of health outcomes and heterogeneous populations of study subjects and practice settings (Tunis, Stryer, and Clancy 2003).

An alternative strategy is to use an unbalanced allocation ratio in which a disproportionately greater number of subjects are randomized to the investigational arm than to the comparison group(s) (Leon and Solomon 2003). A disproportionate allocation ratio (e.g., 2:1, active:comparator) can be used when, for instance, an alternative or usual treatment might be viewed as less desirable by potential subjects. This approach can help with both recruitment and retention.

The question about why a particular treatment works after its efficacy has been established can be answered using "dismantling" or component control designs (Schnurr 2007; Borkovec 1993). In these designs, active ingredients of a treatment believed responsible for its effectiveness are isolated and compared against a form of the treatment that does not contain these elements. A variant of this approach uses an "additive" design that combines therapeutic techniques known to be effective and then compares this combination with the individual techniques alone (Schnurr 2007).

Question 4: MODERATORS AND MEDIATORS. When should tests of potential moderators (subject characteristics associated with treatment responsiveness) and mediators (mechanisms that may explain a treatment effect) be built in?

Suggested Approach: Tests of potential moderators and mediators should be conducted after a treatment is shown to be effective. Exploratory analyses may be used to test mediation or moderation at any stage in the research process (e.g., to determine a treatment's effects across subpopulations), and analyses should be prespecified when possible.

Rationale: A moderator is a variable assessed at baseline that modifies treatment response. Exploratory moderator analyses will focus on the magnitude of the treatment effect for various groups (e.g., between-group effect sizes for those with or without various comorbid disorders such as depression, SUD, and TBI) and should not use significance testing. Results of moderator analyses can be used to guide the design of subsequent RCTs, which strictly recruit particular groups most likely to respond to the intervention.

A mediator is a post-baseline variable that provides information on how a treatment works; i.e., mediators are changes that take place prior to response on the primary outcome. Results of exploratory studies can be used to design future trials but not for making clinical recommendations.

Question 5: DATA ANALYTIC TECHNIQUES. How should primary and secondary data analytic techniques be selected? How should proposed analyses be linked to outcome measure(s)? What standard data analytic plans should be considered to deal with missing data?

Suggested Approach: Data analytic techniques should be selected and prespecified based on the hypothesis, the form of the outcome measure, the projected sample size, and the number of post-baseline assessment time points expected to be included in the primary analyses. The analyses should include data from all randomized subjects, even those with incomplete data. The primary analyses should be conducted according to the principle of intention to treat. That is, each subject's treatment status should be based on the randomized assignment, not on adherence to protocol. Complete case and last observation carried forward, or LOCF, methods of analysis should be avoided and replaced with methods that use all of the available data, such as mixed-effects models for longitudinal data.

Rationale: Data analytic techniques to test a specific hypothesis are chosen based on several characteristics including the form of the dependent variable; e.g., continuous, binary, ordinal, survival. In addition, the sample size and the number of repeated observations per subject have a bearing on the choice of analytic approach. Various methods of including subjects with incomplete data are available. The method to be used should be specified during the design phase, and the sensitivity of this approach should be evaluated and reported.

Mixed-effects models can include all subjects, even those with incomplete data, reducing attrition bias and increasing generalizability, power, and precision (Laird and Ware 1982; Hedeker and Gibbons 2006). Mixed-effects models account for clustering of repeated measures within subject and can account for clustering of subjects within therapist.

Valid inferences from mixed-effects models assume ignorable attrition; i.e., attrition that is accounted for by measures of covariates or the dependent variable that are measured prior to dropout (Laird 1988). One approach to examining this assumption is to ask subjects to rate on Likert scale (at baseline) their "intent to complete" the trial and then at each assessment session to rate their "intent to attend" the next assessment session (Leon, Demirtas, and Hedeker 2007). Another approach to examining this assumption is to implement the pattern mixture model (Little 1993), which can incorporate the pattern of attrition in longitudinal analyses (Hedeker and Gibbons 1997). Furthermore, every effort must be made to continue assessments for the entire course of randomized treatment, even among those who fail to comply with randomized treatment assignment or must leave the study-assigned treatment (Lavori 1992).

Strategies for minimizing missing data should be prespecified and procedures implemented to minimize attrition and incomplete data (Wisniewski et al. 2006). The assessment procedures should be carefully chosen to minimize subject burden and the possibility of attrition. Written documentation should (1) indicate when a subject is considered to have dropped out and (2) distinguish between investigator-initiated protocol deviation and subject-initiated deviation. To reduce dropout, compliance-enhancing interventions should be considered as part of the intervention. This and other efforts to increase compliance should be balanced with generalizability of the outcome. Quality control procedures can assure that forms have been

completed with little incomplete data. Potential quality control procedures might include (1) requiring site personnel to review completed forms before the subject leaves the study site and (2) identifying items that are incomplete and transmitting them to the study sites on a frequent basis. In addition, providing study sites with "report cards" can enhance quality control procedures and thereby strengthen data collection.

Attrition introduces bias and decreases statistical power and generalizability (Leon et al. 2006). If subject attrition is expected to be substantial, post-baseline assessments of outcome should be collected more frequently. A single followup assessment (as in a before-and-after type of design) should not be used. Frequent patient contact can keep subjects engaged in the study, thus reducing the risk of attrition. However, frequent patient contact also increases subject burden and cost. Dropout from treatment should be distinguished from dropout from measurement, and attention should be directed toward how to reduce dropout rates.

Question 6: COMBINED EFFECTS. At what stage in the progression of research on PTSD treatments is it appropriate to study the combined effects of multiple treatments? Under what circumstances does it make sense to conduct studies of the combined effects of psychotherapy and pharmacotherapy for PTSD?

Suggested Approach: In general, the early stages of research should focus on the effectiveness of monotherapeutic interventions before proceeding to multiple therapies. In some instances combined therapies may be appropriate to examine as initial therapies based on the lack of effectiveness of single therapies and theoretical benefits of the combined therapies.

Rationale: An important consideration is whether combined treatments should be prescribed as initial treatment for all patients. Typically they are reserved for patients who are refractory to single treatments alone. In view of cost considerations (for both the patient and facility), it generally makes most sense to determine which single-agent interventions are effective as the first step. Studies that evaluate combined treatments typically require larger recruitment samples relative to studies that evaluate monotherapies.

Suggestions for PTSD Measurement

Question 7: PRIMARY OUTCOME(S). What are the considerations for deciding on the primary outcome measures of clinical and functional status?

Suggested Approach: In selecting the primary outcome measures of clinical and functional status, the following should be considered: (1) Reliability, including interrater reliability, as appropriate; (2) validity, including cultural appropriateness; and (3) practicality. Where resources allow, multiple indicators of the key construct(s) are encouraged.

Rationale: The most important consideration for deciding on the primary outcome measures of clinical and functional status is reliability. If one cannot measure something reliably or consistently, then one cannot expect that measure to be related to something else; i.e., to demonstrate validity. The process of validating a construct (and its measure) is continual. The stronger and longer the evidence for validity, the better. A more reliable assessment process can

decrease sample size requirements (Leon, Marzuk, and Portera 1995) because there is reduced within-group variability. As a result, the size of between-group effects increases and the sample size requirements decrease. The type of reliability depends upon the nature of the measure and the nature of the construct being assessed. If the measure is based on self-report, the measure should, at a minimum, demonstrate internal consistency within its dimensions.

The structure of PTSD is constantly under investigation and is likely more multidimensional than unidimensional, but there is some disagreement on its subdimensions. Research has increasingly encouraged measures of subdimensions of PTSD (Palmieri et al. 2007; King, King, and Orazem 2006; King et al. 1998). It might, therefore, be appropriate that indicators of its dimensions (e.g., intrusion, avoidance, numbing, arousal) be well represented with regard to content saturation and breadth and, ultimately, be highly internally consistent (within each). Measures of PTSD that rely on the 17 Diagnostic and Statistical Manual of Mental Disorders (DSM) symptoms can be scored to reflect indices of dimensions, albeit with limited items per dimension and, hence, limited content representativeness (American Psychiatric Association 2000).

If a PTSD measure, or any measure, is administered by interview with ratings supplied by skilled or semiskilled interviewers, the measure and the raters must demonstrate interrater reliability. If two equally proficient individuals assessing the same person do not arrive at the same conclusion, then the utility of the measure is questionable. In addition, one would require that certain constructs and indicators of those constructs be consistent over assessment occasions.

Another consideration for deciding on the primary outcome measures of clinical and functional status is practicality. One's resources, timeframe, and other constraints on a study have to be balanced against the desire for the more reliable and valid measurement plans. For example, ideally one would have multiple assessments via clinician-administered, structured interviews throughout the treatment regimen and on multiple occasions following the treatment. Very likely, however, this would be difficult to achieve, burdening subjects, the research team, and the funding agency.

Other considerations include the desire for multiple indicators of secondary or exploratory outcomes. Ideally, key outcome variables would be measured from various perspectives; e.g., multiple informants, multiple formats, functioning and satisfaction in multiple domains, multiple settings. If multiple primary outcomes are identified in the protocol, a multiplicity adjustment should be prespecified for hypothesis testing and must be incorporated into sample size determination (Leon 2004). Multiplicity adjustments for highly correlated outcomes can incorporate the correlation in the adjustment (Leon and Heo 2005; Leon et al. 2007).

Question 8: TIMING OF ASSESSMENTS. When should a baseline for PTSD be taken in PTSD treatment trials? What are optimal, critical, or viable time points for psychosocial and physiological assessments of PTSD symptoms relative to (1) the traumatic event, (2) initiation and progress throughout treatment in the course of a clinical trial, and (3) followup post-treatment?

Suggested Approach: Assessments are usually scheduled according to stated expectations about the pace of change during and following a treatment. In an RCT, baseline should be assessed

immediately prior to random assignment. It is critically important to initiate treatment while the baseline score is still an accurate reflection of the person's state prior to treatment. As suggested by the IOM 2008 report, time of assessment for intervention should be based on proposed mechanism of action, technical considerations, baseline measurements, and frequency of measurements. The largest number of assessments should occur around the time when the greatest degree of change is expected. Where resources allow and it can be scientifically justified, longer term posttreatment assessments should be conducted; i.e., greater than 1 year.

Rationale: If possible, obtaining a reliable assessment of symptom course prior to baseline would be valuable. Subjects should be randomized immediately after the baseline assessment. However, feasibility, subject burden, and cost often make multiple baseline observations prohibitively expensive.

With regard to most treatment studies (wherein subjects may enter at varying times since the traumatic event and likely in chronic states), more information is needed about the pool of subjects and the intended treatment to project times of assessments. Little is known about trends over time post-exposure, although research has supported an expected curvilinear concave-downward decrement or negative logarithmic function (King et al. 2003; Koss and Figueredo 2004; Resnick et al. 2007). For the Resnick et al. study, the function is also post-treatment as the participants entered the study and received the intervention just following the time of the traumatic event (rape).

In terms of psychophysiological assessments, the optimal time points depend on a number of factors, including the following:

1) The proposed mechanism of action of the pharmacological or psychotherapeutic intervention; e.g., if propranolol is administered as soon as possible after a traumatic event (during consolidation) or at the time of remembering (reconsolidation), measurement of heart rate, norepinephrine, etc., might be conducted immediately before giving propranolol and then again at least 1 hour after administration of propranolol as well as perhaps weeks or months later.
2) Technological considerations. If one is using structural versus functional imaging to study effects of therapeutic interventions, the timeframe of measurements would be very different. With functional imaging one might measure rapid changes, while with structural imaging proposed changes would be far less rapid and measurements might be spaced further apart.
3) Baseline measurements. Optimal time points for assessment also depend on what the baseline is and whether the research will use active or passive baseline measurements. For some research, baseline could be established with a period of rest followed by a psychophysiological assessment prior to treatment and again after treatment. On the other hand, the research might be focused on stress reactivity so that the baseline measurement would be the change between prestress levels and the highest levels post-stressor administration. Additionally, in some cases diurnal variation would be needed for baseline and followup measurements.
4) Anticipated time of greatest change. Frequency of measurement should depend on when most change is expected to occur. In other words, there should be a greater frequency of measurements during the time of greatest expected change.

To accommodate research subjects, reduce costs, and enhance the ability to gather data, modes of assessment other than inperson assessment should be explored; e.g., Internet, phone, e-mail.

Question 9: LENGTH OF FOLLOWUP. How long should subjects be followed to assess durability of treatment? What issues should be considered in determining adequacy of length of followup and standardization of followup periods?

Suggested Approach: Durability of treatment should be assessed for as long as scientifically justifiable and practically feasible. Timing of assessment should be appropriate to the question being addressed. Separate RCTs might need to be designed to evaluate the acute and maintenance effects of an intervention.

Rationale: In all cases, posttreatment assessment should be conducted with timing appropriate to the question being addressed; e.g., end of last treatment, 1 week, 2 months. Assessing durability of treatment for as long as scientifically justifiable and practically feasible is particularly important (1) when relapse is anticipated, as in the case of military redeployment, and (2) among patients with comorbidities who should undergo longer term observation.

Question 10: MEASURING IMPROVEMENT. How should PTSD treatment improvement be defined? How can response to PTSD treatment (over time and repeated assessments) be measured reliably?

Suggested Approach: A definition of improvement or recovery from PTSD should include, at a minimum, a combination of measures addressing symptom improvement, functional improvement, and quality of life. Measures of recovery relevant to the population should be established and should include decreases in alcohol and drug use for patients with comorbid SUD. Development and evaluation of composite measures should be considered, as should the impact of the phrase "loss of diagnosis" on the patient population in describing recovery from PTSD as it may be perceived to impact health care and/or other benefits. In addition, a definition of recovery from PTSD should include endpoints meaningful not only to therapists and researchers but also to patients and consumers. Nevertheless, a research protocol must prespecify the primary outcome measure; that is, the outcome on which the results of the trial are based. The assessment instrument that is selected should be (1) stable over time in the absence of individual change (test-retest reliability) and (2) sensitive to actual shifts in standing on the attribute. Modes of assessment other than inperson assessment should be explored, and assessment should remain faithful to the research plan even when accommodations are made for subjects' convenience. The use of indicators beyond symptom change to measure improvement or recovery is encouraged.

Rationale: The IOM found "no generally accepted and used definition for recovery in PTSD" and recommended that "clinicians and researchers work toward common outcome measures in three general domains that relate to recovery: loss of PTSD diagnosis, PTSD symptom improvement, and end state functioning" (Institute of Medicine 2008, 149:150). Whereas anecdotally loss of diagnosis would be an ideal test of an intervention, given findings regarding the debilitating nature of subthreshold symptoms (Marshall et al. 2001) it may be more clinically useful to establish measures of high end-state functioning.

Measures of recovery typically have not included alcohol/drug use, which would be of particular relevance for patients with comorbid SUD. Additional questionnaires could include measures of functioning and health-related quality of life; e.g., Short Form Health Survey (SF-36, SF-12), Social Adjustment Scale, Clinical Global Impression of Change, and more specific measures depending on the study.[1] It is important to note that these outcomes reflect therapist or researcher endpoints and may not accurately reflect good end state for a subject.

Question 11: ENDPOINTS. What standard set of primary and secondary endpoints (e.g., activities of daily living, social and occupational functioning) should be considered for all PTSD studies?

Suggested Approach: Depending on study hypotheses/objectives, one of the following should be considered as the primary endpoints in studies that target the reduction of PTSD symptoms as a primary goal: (1) Diagnostic assessment of the presence or absence of PTSD or (2) a continuous index of PTSD symptom severity. Suggested secondary endpoints to be carefully selected and matched to study goals might include (1) comorbid disorders; (2) general physical health functional status and general emotional health functional status; (3) additional measures of social functioning across several specific domains (e.g., occupational/educational, including missed work days, interpersonal conflict, marital/family/parental); (4) indicators of life satisfaction (global or by domain) such as job and marital satisfaction and quality of life; (5) measures of posttraumatic growth; (6) ideographic, client-centered scales of goal attainment; and (7) utilization of health care services.

Rationale: Using the IOM (2008) report as a guide, PTSD studies should include as primary endpoints a diagnostic assessment of the presence or absence of PTSD and a continuous index of PTSD symptom severity keyed to the subdimensions/symptom categories and with stronger content representativeness than existing 17-item self-report measures of the DSM. Studies should also include a set of measures of the core entities that are comorbid with PTSD (e.g., depression, SUD, panic, general anxiety disorder); it might be best to seek continuous indices here as well. Inclusion of conventional measures of general physical health, functional status, and general emotional health functional status can be helpful for calibrating against other studies.

Question 12: MEASURING EXPOSURE TO TRAUMATIC EVENTS. How can exposure to potentially traumatic events and time since exposure be accurately measured? Are there any implications for inclusion in PTSD treatment trials?

Suggested Approach: Self-reporting may be used to measure exposure and time since exposure and should include measures that permit characterization of the trauma experience; e.g., type of trauma, number of traumas, duration of symptoms, time elapsed since trauma, severity of trauma, and chronic illness since exposure. Strategies that improve recall of past events, like the life history calendar method (Axinn, Pearce, and Ghimire 1999), should be considered where

[1] Information about the construction of these instruments are available elsewhere. See Ware, Kosinski, and Keller 1996; Ware et al. 1993, 2000; Weissman and Bothwell 1976; and Guy 1976.

appropriate. In certain situations, objective information or documentation about exposure can be accessed and should be considered.

Rationale: Self-reporting by research subjects is considered a reliable method for measuring exposure to trauma and time since exposure (Dohrenwend et al. 2007). Additional tests of self-report may be helpful. In general, measures of exposure are often based on the most prominent or distressing trauma, although there is appreciation for the challenges around identifying a single trauma among those with PTSD. The moderating effect of *time since exposure* can be evaluated in exploratory analyses as described above (question 4). Considering the difficulties inherent in collecting contextual information related to trauma exposure, self-reporting may allow for greater comparability of results and should be validated if possible.

Question 13: ASSESSMENT MODALITY. How should a trial be designed to achieve a proper balance of self-report and clinician-administered assessments? What are the implications for resources and time/patient burden?

Suggested Approach: The assessment approach should always be tied to the hypothesis. Patient burden as well as assessor burden should be considered when choosing an approach to assessment. For treatment trials, a structured diagnostic assessment interview (Clinician-Administered PTSD Scale [CAPS], PTSD Symptom Scale – Interview [PSS-I], etc.) should be administered and, in most trials, should be used to assess the primary endpoint of PTSD symptoms. Designs requiring multiple observations (e.g., frequent assessments to track the course and trajectory of recovery) can use more easily administered questionnaires; e.g., PTSD Checklist (Weathers et al. 1993).

Rationale: Selection of self-report measures requires the presence of sound psychometric properties of these measures. There is some evidence that symptom reporting may differ from self-report to interview, and this may be more likely in different trauma types. Balance of self-report versus clinician-administered assessment always involves a weighing of subject burden with the research question. For sensitive topics, self-administered questionnaires may be more accurate than interviews (validity issue). The scale of the project also dictates measure selection.

Suggestions for PTSD Trial Implementation

Question 14: RECRUITMENT STRATEGIES. How can feasible and sufficient subject recruitment and enrollment, and adherence to recruitment strategies, be ensured?

Suggested Approach: To ensure feasible and sufficient subject recruitment and enrollment and adherence to recruitment strategies, a variety of strategies relating to (1) sample characteristics, (2) site selection and retention, (3) subject recruitment, and (4) communications should be considered.

Rationale: (1) Sample characteristics. Sample size calculations should be based on reasonable assumptions; e.g., effect size, variability of sample, attrition rates. Prior to enrollment, the following should be evaluated: (a) Availability of the sample and (b) the impact of

inclusion/exclusion criteria. Reasons for exclusion should be routinely examined to ensure that the exclusion criteria remain scientifically sound.

(2) Site selection and retention in multisite trials. It is critical to ensure that there are an adequate number of recruitment sites that have sufficient numbers of patients to meet the target sample size. To account for potential underrecruiting in multisite trials, nonperforming sites can be dropped early in the trial and resources shifted to alternate performing sites where additional capacity is evident. A pool of backup sites with institutional review board approval could be maintained to replace nonperforming sites. Prior to the study, mechanisms should be developed to ensure that performance goals are met and consideration given to placing sites on probation/termination based on poor performance. These plans should be transmitted to the sites so there is an understanding of the requirements for continued funding.

(3) Subject recruitment. Existing databases should be used to identify potential subjects. Adequate incentives can help increase recruitment and retention of subjects in the study. Focus groups can help determine what incentives are most desirable; e.g., travel reimbursement. Subject burden also can be decreased by, for instance, minimizing time needed to collect data and conduct procedures, evaluating recruitment time period and seasonal trends, streamlining the screening and enrollment process, and providing materials to aid in screening; e.g., laminated cards with inclusion/exclusion criteria. Understudied subpopulations (e.g., Reservists and National Guard) should be considered for recruitment, and technologies (e.g., Internet, telephone) that appeal to specific populations (e.g., veterans) should be explored.

(4) Communications. Investigators, study coordinators, and leaders at the clinic or medical center should be engaged via frequent contact and reports, monitoring of goals (e.g., site report cards), and appropriate incentives for achieving goals. Depending on the recruitment mechanism/strategy, it may be appropriate to promote awareness of the study to patients and investigators using, for instance, posters, pamphlets, pens, or notepads.

Question 15: ADHERENCE TO PROTOCOL. How can adherence to protocol be ensured within (and between) sites, over long durations, and by therapists and providers? How can interrater reliability be ensured?

Suggested Approach: Protocol adherence by staff should be encouraged through (1) appropriate training; (2) conducting appropriate site visits; (3) preparing monitoring reports; and (4) taping sessions for adherence, especially for psychosocial interventions. Protocol adherence by patients should be ensured by educating patients about the importance of adherence and monitoring compliance through objective measurement of recommended health behaviors, such as taking prescribed medication or performing homework assignments between therapy sessions. To ensure protocol adherence over long durations, the following strategies are recommend: (1) Review monitoring approaches and dissemination of reports frequently; (2) repeat activities such as training, certification, and reliability on a regular basis (e.g., yearly); and (3) have an independent party conduct fidelity monitoring.

Rationale: To ensure that the treatments being evaluated at the completion of the study are the treatments that the study was designed to evaluate (i.e., intended dose is the same as the dose

received), it is important to assess adherence to the treatment and study protocol over the course of a study. The assessment protocol should include procedures to encourage treatment and protocol compliance, including the conduct of training and certification with periodic (e.g., yearly) retraining and recertification.

Question 16: SUBJECT COMPENSATION. What should be considered in determining subject reimbursement?

Suggested Approach: Subjects should be reimbursed for time, inconvenience, and personal expense incurred as a result of the study visit, outside of time used for standard care treatment. Reimbursement should be sufficient to permit but not coerce participation by disadvantaged populations.

Rationale: The primary cost for patients is time and inconvenience. Psychotherapeutic treatments generally also require a time commitment between sessions (logs, exposure practice, etc.), though reimbursement should be for care that is outside the standard of care. The time cost for patients is a serious consideration, as it relates to their level of commitment and also to the likelihood of continuation of treatment and study participation. Reasonable compensation for patients' time spent and inconvenience is an important component of successful study recruitment and retention, but agency restrictions (e.g., for active duty personnel) on permissible incentives for research participation need to be considered.

Next Steps

Military actions in Vietnam, Afghanistan, and Iraq and numerous natural traumas have focused attention on the mental health needs of and adequacy of services for affected populations. A major emphasis of this attention has centered on the effectiveness of treatments for PTSD. While there was agreement among the workgroup participants that several current treatments are effective for many patients with PTSD, a recent IOM (2008) review of the literature on treatments for PTSD reveals serious challenges to research on the development and testing of various intervention approaches. The approaches suggested here for researchers and expert reviewers on the design, evaluation, and implementation of research are intended to represent a reasonable starting point for strengthening the evidence base for future PTSD interventional studies, and ultimately PTSD treatment.

Research priorities for PTSD have previously been highlighted by the agencies that convened this workgroup (see U.S. Department of Veterans Affairs Office of Research and Development, National Institute of Mental Health, and U.S. Army Medical Research and Materiel Command 2006) and were not further developed in this meeting; however, there was considerable agreement among the participants on the importance of funding high-quality research and the need for improved treatment for PTSD that can best be realized by a concerted, coordinated, and rigorous effort involving the VA and other agencies, including the DoD and the NIMH. The sobering reality is that PTSD affects whole communities, beginning with an individual traumatic exposure and extending into families and society. The overall goal of PTSD research is to provide evidence for the most effective care for and restoration of individuals to their highest level of functioning.

Endnotes

Workgroup Participants

Lori Davis, VA Birmingham/Tuscaloosa; Douglas Delahanty, Department of Psychology, Kent State University; Norah Feeny, Department of Psychology, Case Western Reserve University; Edna Foa, University of Pennsylvania; Thomas Grieger, Defense Center of Excellence of Psychological Health and Traumatic Brain Injury; Col Charles Hoge, Division of Psychiatry and Neuroscience, Walter Reed Army Institute of Research, U.S. Army; Daniel King, Boston University; Lynda King, Boston University; Jerry Larson, Naval Health Research Center; Andrew C. Leon, Weill Cornell Medical College; John March, Duke Child and Family Study Center; Miles McFall, VA Seattle/Puget Sound; Murray Raskind, Puget Sound VAMC, VA Seattle, and Alzheimer's Disease Research Center, University of Washington; Patricia Resick, VA Boston; James Rochon, Duke Clinical Research Institute; Paula Schnurr, National Center for PTSD, VA White River Junction, VT; Steven Southwick, VA West Haven, CT; Stephen Wisniewski, University of Pittsburgh.

Organizing Committee

VA: Terri Gleason, Office of Research & Development, and Peter Peduzzi, Cooperative Studies Program; NIMH: Robert Heinssen, Shoshana Kahana, and Farris Tuma; DoD: Kelley Brix, Office of the Assistant Secretary of Defense for Health Affairs, COL Janet Harris, Medical Research and Materiel Command, U.S. Army; Kenneth Sausen, Center of Excellence, U.S. Navy; Karen Schwab, Walter Reed, U.S. Army.

Acknowledgements

The statements, conclusions, and suggestions contained in this document reflect both individual and collective opinions of the workgroup participants and are not intended to represent the official position of the U.S. Department of Veterans Affairs, the U.S. Department of Health and Human Services, the National Institutes of Health, the National Institute of Mental Health, or the U.S. Department of Defense. Writing and editing assistance in the preparation of this report was provided by Rose Li and Associates, Inc., under contract to the U.S. Department of Veterans Affairs.

References

Altman, D.G., D.F. Schulz, D. Moher, M. Egger, F. Davidoff, D. Elbourne, P. Gotzsche, and T. Lang for the CONSORT Group. 2001. The revised CONSORT statement for reporting randomized trials: Explanation and elaboration. *Ann Intern Med* 134:663–694.

American Psychiatric Association. 2000. *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision.* Washington, DC: American Psychiatric Association.

Axinn, W.G., L.D. Pearce, and D. Ghimire. 1999. Innovations in life history calendar applications. *Soc Sci Res* 28:243–264.

Borkovec, T.D. 1993. Between-group therapy outcome research: Design and methodology. *NIDA Research Monograph* 137:249–289.

Davidson, J.R., D. Hughes, D.G. Blazer, and L.K. Goerge. 1991. Post-traumatic stress disorder in the community: An epidemiological study. *Psychol Med* 21:713–21.

Dohrenwend, B.P., J.B. Turner, N.A. Turse, B.G. Adams, K.C. Koenen, and R. Marshall. 2007. Continuing controversy over the psychological risks of Vietnam for U.S. veterans. *J Trauma Stress* 20(4):449–465.

Glasgow, R.E., D.J. Magid, A. Beck, D. Ritzwoller, and P.A. Estabrooks. 2005. Practical clinical trials for translating research to practice: Design and measurement recommendations. *Med Care* 43:551–557.

Greenwald, P., and J.W. Cullen. 1985. The new emphasis in cancer control. *J Natl Cancer Inst* 74:543–551.

Guy, W. ECDEU Assessment Manual for Psychopharmacology—Revised (DHEW Publ No ADM 76-338). 1976. Rockville, Md.: U.S. Department of Health, Education, and Welfare, Public Health Service, Alcohol, Drug Abuse, and Mental Health Administration, NIMH Psychopharmacology Research Branch, Division of Extramural Research Programs, pp. 218–222.

Hedeker, D., and R.D. Gibbons. 1997. Application of random-effects pattern-mixture models for missing data in longitudinal studies. *Psychol Methods* 2(1):64–78.

Hedeker, D., and R.D. Gibbons. 2006. *Longitudinal Data Analysis*. Hoboken, NJ: John Wiley and Sons.

Institute of Medicine (IOM). 2008. *Treatment of Posttraumatic Stress Disorder: An Assessment of the Evidence*. Washington, DC: The National Academies Press.

Kessler, R.C., W. T. Chiu, O. Demler, and E.E. Walters. 2005. Prevalence, severity and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry* 62:617–627.

King, L.A., D.W. King, and R.J. Orazem. 2006. Research on the latent structure of PTSD. *PTSD Res Q* 17(3):1–3.

King, L.A., D.W. King, D.M. Salgado, and A.Y. Shalev. 2003. Contemporary longitudinal methods for the study of trauma and posttraumatic stress disorder. *CNS Spectr* 38(9):686–692.

King, D.W., G.A. Leskin, L.A. King, and F.W. Weathers. 1998. Confirmatory factor analysis of the Clinician-Administered PTSD Scale: Evidence for the dimensionality of posttraumatic stress disorder. *Psychol Assess* 10(2):90–96.

Koss, M.P., and A.J. Figueredo. 2004. Change in cognitive mediators of rape's impact on psychosocial health across 2 years of recovery. *J Consult Clin Psychol* 72(6):1063–1072.

Laird, N.M. 1988. Missing data in longitudinal studies. *Stat Med* 7(1–2):305–315.

Laird, N.M., and J.H. Ware. 1982. Random-effects models for longitudinal data. *Biometrics* 38(4):963–974.

Lavori, P.W. 1992. Clinical trials in psychiatry: should protocol deviation censor patient data? *Neuropsychopharmacology* 6(1):39–48.

Leon, A.C. 2004. Multiplicity-Adjusted Sample Size Requirements: A Strategy to Maintain Statistical Power when Using the Bonferroni Adjustment. *J Clin Psychiatry* 65:1511–1514.

Leon, A.C., H. Demirtas, and D. Hedeker. 2007. Bias Reduction with an Adjustment for Participants' Intent to Dropout of a Randomized Controlled Clinical Trial. *Clin Trials* 4:540–547.

Leon, A.C., and M. Heo. 2005. A Comparison of Multiplicity Adjustment Strategies for Correlated Binary Endpoints. *J Biopharm Stat* 15:839–855.

Leon, A.C., M. Heo, J.J. Teres, and T. Morikawa. 2007. Statistical Power of Multiplicity Adjustment Strategies for Correlated Binary Endpoints. *Stat Med* 26:1712–1726.

Leon, A.C., C.H. Mallinckrodt, C. Chuang-Stein, D.G. Archibald, G.E. Archer, and K. Chartier. 2006. Attrition in Randomized Controlled Clinical Trials: Methodological Issues in Psychopharmacology. *Biol Psychiatry* 59:1001–1005.

Leon, A.C., P.M. Marzuk, and L. Portera. 1995. More reliable outcome measures can reduce sample size requirements. *Arch Gen Psychiatry* 52:867–871.

Leon, A.C., and D.A. Solomon. 2003. Toward Rapprochement in the Placebo Control Debate: A Calculated Compromise of Power. *Eval Health Prof* 26:404–414.

Little, R.J.A. 1993. Pattern-mixture models for multivariate incomplete data. *J Amer Stat Assoc* 88(421):125–133.

Marshall, R.D., M. Olfson, F. Hellman, C. Blanco, M. Guardino, and E.L. Struening. 2001. Co-morbidity, impairment, and suicidality in subthreshold PTSD. *Am J Psychiatry* 158:1467–1473.

Mercer, S.L., B.J. DeVinney, L.J. Fine, L.W. Green, and D. Dougherty. 2007. Study designs for effectiveness and translation research: Identifying trade-offs. *Am J Prev Med* 33:139–154.

Palmieri, P.A., F.W. Weathers, J. Difede, and D.W. King. 2007. Confirmatory factor analysis of the PTSD checklist and the Clinician-Administered PTSD Scale in disaster workers exposed to the World Trade Center Ground Zero. *J Abnorm Psychol* 116(2):329–341.

Resnick, H., R. Acierno, A.E. Waldrop, L. King, D. King, C. Danielson, K.J. Ruggiero, and D. Kilpatrick. 2007. Randomized controlled evaluation of an early intervention to prevent post-rape psychopathology. *Behav Res Ther* 45(10):2432–2447.

Rounsaville, B.J., K.M. Carroll, and L. Onken. 2001. A stage model of behavioral therapies research: Getting started and moving on from stage 1. *Clin Psychol: Sci Pract* 8:133–142.

Schnurr, P.P. 2007. The rocks and hard places in psychotherapy outcome research. *J Trauma Stress* 20:779–792.

Tunis, S.R., D.B. Stryer, and C.M. Clancy. 2003. Practical clinical trials: Increasing the value of clinical research for decision making in clinical and health policy. *JAMA* 290:1624–1632.

U.S. Department of Veterans Affairs Office of Research and Development, National Institute of Mental Health, and U.S. Army Medical Research and Materiel Command. 2006. Mapping the landscape of deployment related adjustment and mental disorders. A meeting summary of a working group to inform research. http://www.research.va.gov/programs/OIF-OEF/deployment-related-adjustment-may06-report.pdf.

Ware, J.E., M. Kosinski, and S. Keller. 1996. A 12-item short-form health survey: Construction of scales and preliminary tests of reliability and validity. *Med Care* 34(3):220–233.

Ware, J.E., K.K. Snow, M. Kosinski, and B. Gandek. 1993, 2000. *SF-36 Health Survey: Manual and Interpretation Guide.* Lincoln, RI: QualityMetric, Inc.

Weathers, F., B. Litz, D. Herman, J. Huska, and T. Keane. 1993. The PTSD Checklist (PCL): Reliability, Validity, and Diagnostic Utility. Paper presented at the Annual Convention of the International Society for Traumatic Stress Studies, San Antonio, TX

Weissman, M., S. Bothwell. Assessment of social adjustment by patient self-report. 1976. *Arch Gen Psychiatry* 33:1111–1115.

Wisniewski, S.R., A.C. Leon, M.W. Otto, and M.H. Trivedi. 2006. Prevention of Missing Data in Clinical Research Studies. *Biol Psychiatry* 59:997–1000.

070108