Published in final edited form as:

Acad Psychiatry. 2009 ; 33(3): 221–228. doi:10.1176/appi.ap.33.3.221.

Integrating Statistical and Clinical Research Elements in Intervention-Related Grant Applications: Summary from an NIMH Workshop

Joel T. Sherrill, Ph.D.,

Division of Services and Intervention Research (DSIR) at NIMH

David I. Sommers, Ph.D.,

Division of Extramural Activities at the NIMH

Andrew A. Nierenberg, M.D.,

Department of Psychiatry at MGH

Andrew C. Leon, Ph.D.,

Department of Psychiatry at Weill Medical College of Cornell University

Stephan Arndt, Ph.D.,

Iowa Consortium for Substance Abuse Research at University of Iowa Hospitals and Clinics

Karen Bandeen-Roche, Ph.D.,

Department of Biostatistics at Johns Hopkins Bloomberg School of Public Health

Joel Greenhouse, Ph.D.,

Department of Statistics at Carnegie Mellon

Donald Guthrie, Ph.D.,

Professor Emeritus of Biostatistics at UCLA

Sharon-Lise Normand, Ph.D.,

Harvard Medical School

Katharine A. Phillips, Ph.D.,

Department of Psychiatry & Human Behavior at Brown University

M. Katherine Shear, M.D., and

School of Social Work at Columbia University

Robert Woolson, Ph.D.

Department of Psychiatry/Biostatistics at Medical University of South Carolina

Abstract

Objective—To summarize points for consideration generated in an NIMH workshop convened to provide an opportunity for reviewers from different disciplines – specifically clinical researchers and statisticians – to discuss how their differing and complementary expertise can be well integrated in the review of intervention related grant applications.

Address correspondence to David I. Sommers, National Institute of Mental Health, Division of Extramural Activities, 6001 Executive Boulevard, Room 6144, Bethesda, MD 20892; dsommers@mail.nih.gov.

Views expressed within this article represent those of the authors and are not intended to represent the position of NIMH, NIH, or DHHS.

Methods—A one day workshop was convened in October, 2004. The workshop featured panel presentations on key topics followed by interactive discussion. This manuscript summarizes the workshop and subsequent discussions, which centered on topics including: weighting the statistics/data analysis elements of an application in the assessment of the application's overall merit; the level of statistical sophistication appropriate to different stages of research and for different funding mechanisms; some key considerations in the design and analysis portions of applications; appropriate statistical methods for addressing essential questions posed by an application; and the role of the statistician in the application's development, study conduct, and interpretation and dissemination of results.

Results—A number of key elements crucial to the construction and review of grant applications were identified. It was acknowledged that intervention related studies unavoidably involve trade-offs. Reviewers are helped when applications acknowledge such trade-offs and provide good rationale for choices made. Clear linkage among the design, aims, hypotheses and data analysis plan and avoidance of disconnections among these elements also strengthens applications.

Conclusions—Multiple points to consider when constructing intervention related grant applications were identified. The points are presented here as questions and do not reflect institute policy or comprise a list of best practices, but rather represent points for consideration.

Introduction

This report summarizes a workshop that the NIMH held on October 12, 2004, focused on statistical and methodological issues that often appear in the review of interventions research grant applications. Reviews are performed by a multidisciplinary group of reviewers that includes doctoral level clinical researchers who are familiar with the disorders and their treatments, as well as statisticians. Additional scientific experts and public participant reviewers, who represent various stakeholder perspectives are recruited to ensure that each application gets a comprehensive, fair, expert, and balanced review.

Diverse perspectives in the review process add breadth and depth to the review, but statistical reviewers and clinical reviewers frequently have different perspectives that can add to the complexity of peer review. In addition, the nature of contemporary clinical research also adds to the complexity of the review process. Challenges in designing and conducting intervention studies include, for example: defining inclusion and exclusion criteria to select the target population; identifying promising and innovative treatments; selecting appropriate comparison conditions and designs that yield maximally interpretable, valid, and clinically useful results; implementing reliable, valid, and well-timed assessments to characterize intervention processes and outcomes; ensuring that the interventions are delivered with fidelity; and selecting and applying appropriate data-analytic techniques that will illuminate intervention effects and moderators/mediators of response. Issues involving the appropriateness of the proposed study's aims, the design, and the planned data analysis, as well as the precision of fit among these elements of the application, are central issues within the review process.

The review process can seem mysterious to applicants who have not experienced it either in the form of receiving feedback from a review or participating in a review meeting. Thus, the purpose of this paper is threefold:

1. to summarize points for consideration that were generated during the workshop and in subsequent discussions so that researchers who submit interventions applications can understand some of the design- and statistics-related components of review and submit applications with a better understanding of the expectations of both clinical and statistical reviewers.

2. to de-mystify and make transparent some, but certainly not all, of the core issues that show up in review meetings where the clinical and statistical elements of an application are judged and weighed by reviewers who approach the review process with different perspectives
3. to hopefully also serve as an enduring reference for future reviewers of intervention research applications.

Workshop Summary

The workshop summary is structured as a series of “points for consideration” formatted as questions, rather than as a list of explicit recommendations. The points reflect the participants’ opinions that the optimal content and level of detail for a given application depend on its research focus and scope of the proposed work. It should be emphasized that the workshop focused on issues relevant to the review of mental health *interventions research applications* and that the following points for consideration will not apply equally to all types of grant applications. The following points for consideration are intended as a potential resource for those who develop mental health interventions research applications and those who review them.

Weighing the Design and Analysis Elements of an Application in the Assessment of the Application’s Overall Merit

A central impetus for convening the workshop was to discuss how clinical research elements and design/analytic elements of applications should be considered and weighed in peer review. A clinically meaningful research question *and* sound methods and analyses are both necessary components of competitive applications. At times, however, there is a disconnect between or imbalance in the quality of these components. Extreme examples might involve compelling research questions paired with mismatched, under-developed, or seriously flawed analysis plans. Conversely, elegant analytic plans may be applied to less significant research questions.

At times, the review process identifies serious statistical and methodological problems, sometimes characterized by reviewers as “fatal flaws” that threaten the study’s ability to address the primary research question. These problems include but are not limited to selection effects, insufficient power in later-stage intervention studies, and differential attrition or other problematic sources of missing data that threaten the interpretability of the results. Many concerns identified in reviews represent more minor weaknesses that merely detract from the merit of an otherwise potentially viable application. In such circumstances, the overall evaluation of merit ultimately depends on the degree to which the application, in its current form, promises to yield meaningful results; the degree to which adjustments (e.g., different analytic strategies) could improve the application; and the degree to which the investigator team appears capable of addressing problems. Discussion regarding weighing the elements of the application of the application is summarized in the following points for consideration.

- Are the study design and plans for assessment and data-analysis optimal for addressing the study questions, and is the overall plan likely to yield interpretable data? For later-stage intervention testing and larger trials, are the plans supported by preliminary data? For earlier-stage efforts where pilot data per se are not required: does the application provide evidence from the literature or from preliminary studies to support the general idea underlying the intervention, supply evidence to support the feasibility of conducting the research (e.g., capacity to identify and enroll participants), and address how the data will usefully inform a larger, more definitive trial?

- If a flaw in the proposed statistical analysis is identified, can it be fixed, and does the investigative team have the ability to do so?
- In cases where more than one data analytic strategy might be applied, is the chosen strategy well defended, and is the rationale for ruling out other strategies clearly presented? Are limitations and compensatory strategies discussed?

Expectations Regarding the Statistical Elements for Various Funding Mechanisms

Some studies are primarily aimed at advancing our understanding about the efficacy or effectiveness of an intervention and/or identifying moderators and mediators of response. Other studies that occur earlier in the research process are primarily focused on intervention development or novel adaptations of existing interventions and thus constitute pilot studies, where the goal is to assess feasibility, acceptability, safety and tolerability (1). Depending on the type of study, and accordingly, the funding mechanism, the nature of the statistical and methodological issues will vary (information about specific funding mechanisms is available at www.nimh.nih.gov). Relevant points for consideration are summarized, below.

- Will the nature and volume of the data support the proposed analyses in a manner that does not violate basic assumptions (e.g., assumptions regarding distributions)?
- Are research questions and hypotheses prioritized (e.g., as primary, secondary, exploratory) and are analyses and power assumptions addressed accordingly? For example, in addition to addressing primary questions regarding patient-level outcomes, larger efficacy studies often address secondary (non-symptom) outcomes and/or test hypotheses regarding moderators and mediators of response. Similarly, effectiveness trials may address secondary questions regarding how patient-, provider-, and system-level factors affect therapists' implementation of treatment.
- For preliminary or pilot studies, are the aims discussed appropriately (e.g., in terms of feasibility, acceptability), and do the data analyses match the aims? In cases where a study will not be fully powered, is there a well-articulated plan for collecting information that will inform later, more definitive tests of the intervention? Pilot studies can yield important information regarding the feasibility of: identifying, enrolling, randomizing, and retaining participants; training evaluators and therapists; and administering the assessment and intervention protocols (2) (see additional related discussion under the section on "power analysis").
- For later-stage studies and larger projects, is information from earlier work supplied to demonstrate both the feasibility of conducting the larger study *and* the capacity to answer the research question and detect a clinically meaningful treatment effect?

Considerations in the Review of Design and Analysis Elements of Applications

Much of the workshop discussion addressed aspects of the design and analysis portions of applications that impact whether the proposed study will be sufficient or optimal for addressing the research questions. Discussion focused on general considerations as well as specific topics, including strategies for minimizing and handling missing data, effect size considerations, and power analyses.

General Considerations

- Is there clear linkage (a path) between the study aims, hypotheses, data-collection plan, and proposed analyses? The importance of a direct correspondence between the research questions and the study design and analysis was a fundamental theme in the workshop and subsequent discussions. Disconnects in these important elements are commonly cited problems. Workshop participants agreed that a lack of

correspondence among key components of the application can significantly undermine the merit of an application that otherwise addresses an interesting and important research question.

- Are the hypotheses testable/falsifiable and clearly articulated? Would the results be of interest regardless of the direction of results?
- Are all of the instruments and assessment points incorporated in the analyses? Are multiple outcomes and secondary hypotheses (e.g., moderator/mediator analyses) addressed in the data analysis section? Are measures included in lists or tables (or in the appendices) and addressed in the data analysis section? The inclusion of redundant instruments and excessive administration of assessments that may never be analyzed can unnecessarily increase respondent burden and detract from an application's merit.
- Are the strategies for allocating participants to study conditions described in sufficient detail (e.g., randomization scheme, unit of randomization, stratification, balance)? Are operational details regarding how, when, and by whom randomization will be conducted provided?
- Is the rationale for the selection of the comparison condition (e.g., placebo, active comparison, usual care) provided? The main challenge is to make a cogent argument supporting the selection of comparison condition(s) and the rationale for ruling out alternatives strategies (3,4,5,6).
- Are methods to protect data integrity (e.g., provisions for thorough data entry, security, and transport) and data-checking methods (e.g., strategies for detecting and addressing outliers, unusual distributions, and missing data) adequately detailed? As noted in the report of the NIMH task force for mental health intervention research (7), the operational details of data management and oversight are critical elements of treatment research.

Missing Data

Missing data occur for a variety of reasons and affect both outcomes and covariates: The main consequence of missing data is potential bias in the estimate of the treatment effect; specifically, subjects for whom there is “completely” observed information may be systematically different from subjects with incompletely observed information. The workshop discussion acknowledged two fundamental facts regarding missing data: all studies are at risk for missing data and, moreover, the potential for bias arising from missing data exists even if the investigator proposes to undertake an intention-to-treat analysis. The following related points for consideration were noted.

- Are methods to minimize missing data discussed (8)? Are retention/attrition-prevention strategies and strategies to maximize complete data from subject evaluations described in sufficient detail? When participants decline further treatment but are willing to participate in study assessments, are plans specified for continued follow-up and for handling data from those who terminate treatment prematurely?
- Is there appropriate attention to missing data, and are estimates of the amount of anticipated missing data justified and realistic?
- Is the mechanism by which data may be missing considered (e.g., missing item responses, difficulty contacting participants, illness, dropout from the study)? Are these estimates and suspected mechanisms of missing data compatible with statistical assumptions (e.g., are data likely to be missing at random)?

- Are appropriate strategies for handling missing data presented in sufficient detail (e.g., imputation, multiple imputation, likelihood model-based approaches), and are problematic strategies avoided (e.g., complete case analyses) (9,10)?

Effect Size Considerations

An effect size is an index that measures the size of a treatment effect and is independent of sample size (11): Effect sizes are key components of treatment related applications, as they may be used to represent the smallest effects that would be clinically meaningful.

- Is the case for clinical significance or clinical meaningfulness considered? Or, rather, are estimated effect sizes from pilot studies used without clinical interpretation? While there is little consensus regarding the external standards that should be used to define clinical significance (e.g., in terms of normative functioning, symptom diminution, remission of diagnoses), and there are no formal tests of “clinical significance,” various effect size measures that can facilitate interpretation of clinical significance have been proposed (12).
- Is the targeted effect size discussed with reference to a clinically meaningful difference in terms of units of *the primary outcome measure*?

Power Analyses—There are many aspects to a power analysis, but at the heart of any discussion of statistical power, a grant reviewer expects to find the explicit quantitative assumptions that are relevant to whether the proposed study design can answer the research question(s) of interest. A power analysis requires the explicit articulation of the specific outcome of interest and how it is quantified. Inherent in power considerations are a number of key questions: What is a clinically meaningful effect? What is the relation between the primary outcome measure and the research question of interest? How variable is the outcome of interest?

- Do the power calculations and sample size estimates take into account detection of a clinically meaningful effect? Discussions of power and effect sizes in clinical trials emphasize the importance of identifying a threshold of clinical significance, such that any effect size below this threshold difference would indicate that the difference between treatment conditions (e.g., the experimental treatment versus placebo) would not be clinically meaningful (11,13). Nonetheless, investigators sometimes base determinations regarding adequate power on the basis of conventions or desirable effect sizes (e.g., a “medium” effect). Alternatively, power calculations are sometimes based on effect sizes observed in pilot studies. Various authors have cautioned against the use of effect sizes from prior studies for projecting power and calculating sample sizes (2,14). Kraemer et al. (2) cogently illustrate the danger of using previously-observed effect sizes for projections when the standard error of an effect size is large, as is typical in small pilot studies, and caution against the use of data from small studies to guide projections for applications for larger research trials. Instead, these authors emphasize the importance of pilot studies for demonstrating feasibility. Additional data that describe the identification, enrollment, randomization, and retention of participants; training of evaluators and therapists; and use of the measurement and intervention protocols may also be usefully collected and examined in pilot work. Additionally, attention to and explication of the observed variability in how participants respond to the intervention in a pilot study can be useful. Discussion at the workshop noted measures such as confidence intervals around estimated effect sizes and number needed to treat indices (i.e., the number needed to treat to prevent one adverse outcome) (15) might be useful in this regard.
- For small projects proposed in career development applications or in exploratory or developmental pilot studies, does the application clearly acknowledge power

limitations in the context of the study goals (e.g., preliminary feasibility study) and/or budget limitations?

Does the description of the power analysis reference the necessary components: hypothesis, specific data-analytic strategy employed, type I error, power, population effect size, and sample size? Is power addressed for each of the primary hypotheses that are proposed, and does the presentation of the power analysis for each hypothesis address the four essential interdependent components of power analyses -- type I error rate (α), power, population effect size, and sample size? Given that the results of the proposed study are unknown, a display of a range of effect sizes with corresponding sample sizes can be useful to estimate the sensitivity of the power calculations to effect size assumptions.

- Are multiple comparisons considered in the type I error rate? A type I error rate of $\alpha=.05$ is the general convention, although certain situations might suggest use of other values. If multiple primary outcomes are specified in the application, α adjustments for multiple testing should be considered, with the power analyses incorporating an adjusted alpha-level, and multiplicity-adjusted sample size estimates may be appropriate in estimating the necessary sample size (16). The resulting sample size that is necessary will increase by about 20% with two primary endpoints and by 30% with three primary endpoints. As a result, there will be corresponding increases in research costs and study duration. Thus, competitive applications tend to use multiple endpoints judiciously when necessary to answer the study questions.
- Is dependence of observations considered? A common pitfall is the failure to account for dependence among observations, which is relevant in group-based interventions and in community or practice-based research (e.g., in school-based interventions research where participants are nested within classrooms within schools).
- Are two-tailed tests used? Is the directionality of the hypothesis tests (one-tailed vs. two-tailed tests) clearly addressed, and if two-tailed tests are *not* being used, is an appropriate justification provided for one-tailed tests? It is often critically important to detect both the hypothesized beneficial and unanticipated, detrimental effects of the intervention. This typically necessitates the use of two-tailed tests.
- Do power analyses correspond to the statistical procedure that is proposed for addressing the hypotheses? When generating sample size algorithms for less familiar statistical procedures, are appropriate justifications, references, and descriptions of power analysis software provided?

It was also noted that in addition to the *scientific interest* in the power analysis and sample size calculations, there is an *ethical imperative* associated with these analyses that should be considered, along with a host of other ethics issues in psychiatric research (17). Researchers have an ethical obligation to carefully consider the sample size proposed for a research study (18). In fact, the Committee on Professional Ethics of the American Statistical Association issued Ethical Guidelines for Statistical Practice states, "Avoid the use of excessive or inadequate numbers of research subjects by making informed recommendations for study size" (19). Those *informed recommendations* are based on statistical power analyses, which integrate information from specific aims, hypotheses, preliminary studies, and methods (i.e., subjects, assessments, and data analytic procedures).

Appropriate (and Inappropriate) Statistical Methods for Addressing the Essential Questions Posed by an Application

A key consideration in the evaluation of an application is whether the proposed statistical methods are appropriate in terms of the specific research questions and in terms of the data

that will be collected (e.g., the amount of data, the nature and anticipated distribution of the data). Another consideration is the appropriate degree of complexity: competitive applications appropriately exploit modern analytic methods to yield maximally interpretable results and reveal robust effects. Perhaps ideally, the analyses would utilize the simplest methods likely to yield valid, precise, meaningful, and appropriately targeted results. The following points for consideration are potentially useful for avoiding methods that may be overly simple or strategies that entail considerably greater complexity than appears justifiable given the study design, aims, and data.

- Are distributions of the data to be analyzed, descriptive statistics and modeling assumptions underlying any planned analyses specified?
- Are complex data-analytic methods (e.g., highly parametric methods such as hierarchical models, structural equation modeling) employed only when appropriate? Specifically, does the application outline preliminary analyses or a plan to validate that modeling assumptions are reasonably met in the data to be analyzed?
- When complex methods are proposed, is there a convincing argument that the proposed modeling complexity is needed to achieve inferences that are valid, precise, and directly address essential questions?

The Role of the Statistician in the Application's Development, Study Conduct and Interpretation and Dissemination of Results

The integration of the clinical research and design and analyses aspects of interventions research is embodied in the collaboration between the clinical researchers and statistician collaborators who form the research team. The depth and quality of this collaboration is often evident, not just in the data analytic section of the application, but throughout (e.g., in the framing of the aims, in the data collection plan, and even in the budget justification). The following points for consideration reflect discussion regarding the statistician's role.

- Are the roles of the clinical investigators and the statistician structured as collaboration or as a division of labor? (20,21)? Are the statisticians' specific roles and responsibilities clearly outlined (e.g., as consultant, project statistician, co-investigator, core director)? Depending on the nature and complexity of the design and analytic plan, multiple levels of statistical support might be appropriate. For example, on-site statistician collaborators might assume responsibility for interfacing with clinical investigators and executing the analyses, while additional experts might be consulted about less commonly used techniques or specific aspects of the data analytic plan.
- Do letters of support coincide with the roles described in the body of the application and in the budget justification? Detailed, consistent information regarding statistician roles (e.g., across the body of the application, the budget sections, and the support letters) and realistic estimates of the statisticians' effort to accomplish the work provide assurance that the collaboration has been considered thoughtfully.
- Is the statistician qualified as evidenced by training, academic standing, experience, and record of publication/funding? In competitive applications, the biosketches and preliminary studies sections (if relevant) of the application provide evidence that the statistician has experience/familiarity with the proposed data-analytic methods and has a history of working productively with the investigators or with other research teams.
- Is the statistician's input evident *throughout* the application in: the framing of the hypotheses, the data-collection plan, the data-management plan, and the data-analysis plan? While the topic of the biostatistician's role in clinical research has received

relatively limited explicit attention (but see 22,21), it is generally agreed that the statistician's input is crucial, not just in the authoring of the data analysis plan, but throughout the development of the research application.

- Is there adequate statistical support/effort for: monitoring data integrity, executing analyses, assisting with interpretation of the results, and disseminating results? Given the increasingly complex nature of interventions research, consistent attention to the "operations" elements of trials is increasingly important (7). In strong applications, the statistician's role prior to the analyses (e.g., in overseeing data quality/integrity, in preparing analytic ready datasets) and after conducting the analyses (e.g., in interpreting results and reconciling them with prior findings, in producing reports for dissemination of results) is comprehensively addressed.
- For Mentored Career Development applications, is the biostatistician Consultant/Mentor's input evident: in the data-analytic plan for the research project and in the selection of the coursework/training activities? Is the proposed statistical coursework appropriate to the candidate's prior training, matched to the specific training objectives, and relevant to interventions research? Most clinical researchers who submit mentored career development applications propose additional training in methods and statistics as a key part of their career development plan. In this context, the statistician's role extends beyond that of the typical research application and includes a role as a mentor. Accordingly, the application and the corresponding support letter from the statistician consultant/mentor should detail the purpose and frequency of meetings with the trainee and should elaborate how these activities will contribute to the candidate's proposed research project and goal of becoming an independent investigator.

Conclusion

The complex nature of contemporary clinical research necessitates bringing multiple viewpoints to bear in both the development and execution of research studies, and in the review process during which the application's merit is evaluated. The purpose of the workshop and this summary was to address the common issues at the interface of clinical research and design/analytic elements of studies that pose challenges to researchers and to reviewers. Meaningful collaboration is clearly needed between experts who are familiar with the clinical elements of a disorder, existing treatments, and the needs of the health care field, and experts in design and analyses relevant to clinical trials research. This partnership is necessary to properly address clinically meaningful questions and identify studies that will yield interpretable results that can lead to improved treatment of mental and behavioral disorders.

Acknowledgments

We thank the extramural scientists, public-participant reviewers, and NIMH Review and Program Staff who participated in this NIMH-sponsored workshop for their invaluable input. The authors also acknowledge the ongoing contributions of those who serve in the peer review process; their feedback was instrumental in conceptualizing the workshop and in the preparation of this report.

References

1. Rounsaville BJ, Carroll KM, Onken LS. A stage model of behavioral therapies research: Getting started and moving on from stage I. *Clinical Psychology: Science and Practice* 2001;8(2):133–142.
2. Kraemer HC, Mintz J, Noda A, Tinklenberg J, Yesavage JA. Caution regarding the use of pilot studies to guide power calculations for study proposals. *Arch Gen Psychiatry* 2006;63:484–489. [PubMed: 16651505]

3. Appelbaum PS, Roth LH, Lidz CW, Benson P, Winslade W. False hopes and best data: Consent to research and the therapeutic misconception. *Hastings Center Report* 1987;20–24. [PubMed: 3294743]
4. Leber PD. The hazards of inference: The active control investigation. *Epilepsia* 1986;30(Suppl 1):S57–S63. [PubMed: 2673757]
5. Leon AC, Solomon DA. Toward Rapprochement in the Placebo Control Debate: A Calculated Compromise of Power. *Evaluation and Health Professions* 2003;26:404–14.
6. Temple R, Ellenberg SS. Placebo-controlled trials and active-control trials in the evaluation of new treatments. Part 1: Ethical and scientific issues. *Ann Intern Med* 2000;133:455–63. [PubMed: 10975964]
7. National Advisory Mental Health Council. Treatment Research in Mental Illness: Improving the Nation's Public Mental Health Care through NIMH Funded Interventions Research. A Report by the National Advisory Mental Health Council's Workgroup on Clinical Trials. 2005. (Available at http://www.nimh.nih.gov/council/interventions_research.cfm)
8. Wisniewski SR, Leon AC, Otto MW, Trivedi MH. Prevention of Missing Data in Clinical Research Studies. *Biol Psychiatry* 2006;59 (11):997–1000. [PubMed: 16566901]
9. Beunckens C, Molenberghs G, Kenward MG. Direct likelihood analysis versus simple forms of imputation for missing data in randomized clinical trials. *Clin Trials* 2005;2(5):379–386. [PubMed: 16315646]
10. Leon AC, Mallinckrodt CH, Chuang-Stein C, Archibald DG, Archer GE, Chartier K. Attrition in Randomized Controlled Clinical Trials: Methodological Issues in Psychopharmacology. *Biol Psychiatry* 2006;59 (11):1001–1005. [PubMed: 16503329]
11. Cohen, J. Statistical Power Analysis for the Behavioral Sciences. 2. Hillsdale, NJ: Lawrence Earlbaum Associates; 1988.
12. Kraemer HC, Morgan GA, Leech NL, Gliner JA, Vaske JJ, Harmon MD. Measures of clinical significance. *J Am Acad Child Adolesc Psychiatry* 2003;42(12):1524–1529. [PubMed: 14627890]
13. Kraemer HC, Kupfer DJ. Size of treatment effects and their importance to clinical research and practice. *Biol Psychiatry* 2006;59:990–996. [PubMed: 16368078]
14. Korn EL. Projections from previous studies: a caution. *Controlled Clinical Trials* 1990;11:67–69. [PubMed: 2157583]
15. Cook RJ, Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. *BMJ* 1995;310:452–454. [PubMed: 7873954]
16. Leon AC. Multiplicity-Adjusted Sample Size Requirements: A Strategy to Maintain Statistical Power when Using the Bonferroni Adjustment. *J Clin Psychiatry* 2004;65:1511–1514. [PubMed: 15554764]
17. Roberts LW, Solomon Z, Roberts BB, Keith SJ. Ethics in psychiatry research. *Acad Psychiatry* 1998;22:1–20.
18. Halpern SD, Karlawish JHT, Berlin JA. The continuing unethical conduct of underpowered clinical trials. *JAMA* 2002;288(3):358–362. [PubMed: 12117401]
19. American Statistical Association Committee on Professional Ethics: Ethical guidelines for statistical practice. 1999. (Available at <http://www.amstat.org/>)
20. Tilley, A. An Introduction to Research Methodology and Report Writing in Psychology. Brisbane: Pineapple Press; 1999.
21. Ellenberg JH. Biostatistical Collaboration in Medical Research. *Biometrics* 1990;46:1–32. [PubMed: 2190637]
22. Bailar, JC, III. Communicating about statistics with a scientific audience. In: Bailar, JC., III; Mosteller, F., editors. *Medical Use of Statistics*. NEJM Books; 1986.