Special Article

Psychotherapy and Psychosomatics

Psychother Psychosom 2009;78:275–284
DOI: 10.1159/000228248

Received: June 17, 2008
Accepted after revision: October 31, 2008
Published online: July 11, 2009

The Selection and Design of Control Conditions for Randomized Controlled Trials of Psychological Interventions

David C. Mohr[a, b] Bonnie Spring[a, b] Kenneth E. Freedland[c] Victoria Beckner[d] Patricia Arean[d] Steven D. Hollon[e] Judith Ockene[f] Robert Kaplan[g]

[a]Department of Preventive Medicine, Northwestern University, Chicago, Ill., [b]Hines VA Medical Center, Hines, Ill., [c]Department of Psychiatry, Washington University School of Medicine, St. Louis, Mo., [d]Department of Psychiatry, University of California, San Francisco, Calif., [e]Department of Psychology, Vanderbilt University, Nashville, Tenn., [f]University of Massachusetts Medical School, Worchester, Mass., and [g]School of Public Health, University of California, Los Angeles, Calif., USA

Key Words
Randomized controlled trial, methodology · Randomized controlled trial, control conditions · Psychological interventions

Abstract
The randomized controlled trial (RCT) provides critical support for evidence-based practice using psychological interventions. The control condition is the principal method of removing the influence of unwanted variables in RCTs. There is little agreement or consistency in the design and construction of control conditions. Because control conditions have variable effects, the results of RCTs can depend as much on control condition selection as on the experimental intervention. The aim of this paper is to present a framework for the selection and design of control conditions for these trials. Threats to internal validity arising from modern RCT methodology are reviewed and reconsidered. The strengths and weaknesses of several categories of control conditions are examined, including the ones that are under experimental control, the ones that are under the control of clinical service providers, and no-treatment controls. Considerations in the selection of control conditions are discussed and several recommendations are proposed. The aim of this paper is to begin to define principles by which control conditions can be selected or developed in a manner that can assist both investigators and grant reviewers.

The randomized controlled trial (RCT) provides critical evidence for the efficacy or effectiveness of psychological interventions [1]. In an RCT, the efficacy or effectiveness of an experimental treatment is always determined relative to a control condition. Consequently, what an RCT reveals about the effectiveness of the experimental treatment inherently depends as much on the control condition as on the experimental treatment.

There is considerable heterogeneity in the forms of control conditions used for RCTs of psychological interventions, and little agreement or consistency in how to design or select them. Different types of control conditions produce significantly different effects on the outcomes of RCTs [2]; thus, the choice of control conditions has a substantial impact on the evidence that underpins evidence-based practice. Most of the literature on control conditions has focused on general descriptions of meth-

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

0033–3190/09/0785–0275$26.00/0

Accessible online at:
www.karger.com/pps

David C. Mohr, PhD
Department of Preventive Medicine, Northwestern University
680 N. Lakeshore Dr., Suite 1220
Chicago, IL 60611 (USA)
Tel. +1 312 503 1403, E-Mail d-mohr@northwestern.edu

odology [3] or on debating the merits of a particular type of control condition (e.g. placebo controls) [4, 5]. In contrast, the aim of this paper is to begin to define principles by which control conditions can be selected or developed in a manner that can assist both investigators and grant reviewers. To this end, we will review the role of control conditions in eliminating alternative explanations for RCT outcomes, examine three major classes of control conditions, discuss factors to be considered in control group selection, and make suggestions for next steps towards creating clearer and more systematic guidelines for control condition selection and development.

Why Control Conditions Are Needed

The use of control conditions and randomization are the principal methods ensuring that an RCT is 'internally valid'. Internal validity refers to the extent that the outcome for a trial can be attributed to the experimental treatment and not to any alternative explanation such as the natural course of the target problem. The original conceptualization of threats to validity articulated by Campbell and Stanley [6] remains largely unchanged today [7], but the RCT design has evolved considerably. Paradoxically, some contemporary RCT procedures designed to address threats to internal validity may create new problems. Below, we review three threats to internal validity that arise from modern RCT methodology. These threats have not been adequately considered to date and can be aggravated by inadequate control condition design.

Treatment Fidelity Procedure Effects

Several procedures designed to ensure that experimental and control treatments are delivered as intended are now widely accepted by clinical trialists [8]. The treatment should be manualized, its implementation monitored, and the interventionists trained and supervised. These procedures have been shown to improve intervention quality, reduce variability among clinicians, and improve trial outcomes [9]. However, they are often applied differentially across treatment arms. For example, no-treatment, attention, and treatment-as-usual (TAU) control conditions often lack monitoring and feedback protocols, and are less rigorously defined than experimental treatments. If treatment fidelity monitoring procedures are administered with different levels of specificity, rigor, or enthusiasm across treatment arms, internal validity may be threatened.

Clinician Selection and Allegiance Effects

The skills, abilities, and attitudes that treating clinicians bring to the intervention may influence the outcomes of an RCT. Rigorous clinician training is recommended, but most of the clinical skills and abilities that study clinicians utilize are acquired long before they begin work on a trial. Treating clinicians are usually selected for their training, background, and familiarity with the treatment modality or with the theoretical orientation instantiated in the experimental treatment [10]. Clinicians also tend to arrive with an allegiance to the treatment they are to deliver, or develop one during the trial. If clinicians provide a treatment that they are well trained in, believe in, or know to be the experimental treatment, they will likely be more skilled and enthusiastic in their delivery of that treatment, which in turn may influence the outcome [11]. Failure to maintain equivalence in therapist selection and allegiance may threaten internal validity.

Control Condition Effects

Although control conditions are an essential method of managing threats to internal validity, they may also have unintended effects on the outcomes of RCTs. Knowledge that one is not receiving treatment affects outcomes. Patients who become unblinded and learn they are receiving placebo do more poorly than patients who do not know they are receiving placebo [12, 13]. Participants randomized to wait-list control (WLC) conditions have been noted to improve less than would be expected in an equivalent sample studied observationally [14, 15]. Asking WLC subjects to refrain from seeking treatment may decrease natural help-seeking behaviors. Assignment to 'no treatment' may strengthen participants' beliefs that they will not improve, thereby reducing the chance of spontaneous improvement [16]. Some control conditions may also be less credible than the experimental treatments with which they are paired. Treatment credibility has been shown to influence outcomes, particularly for brief interventions [17]. In short, control conditions may create threats to internal validity by influencing outcome expectancies and/or altering health-promoting behaviors.

Commonly Used Control Conditions

Control conditions can be divided into three classes: (a) those in which the treatment is defined by and under the control of the investigator, (b) those in which the

treatment is not defined by the investigator or provided by the study, and (c) no-treatment control conditions. We will review each of these classes of control conditions with attention to the potential benefits and threats to internal validity. This discussion is summarized in table 1.

Control Conditions Composed of a Study-Defined Treatment

Investigators have the greatest influence over control conditions when they define and implement them. These conditions include both specific and nonspecific treatment component controls.

Specific Treatment Component Controls

Some control conditions provide one or more specific components of the active treatment, thereby creating an additive or a dismantling design. They make it possible to determine whether the additional treatment components present in the experimental treatment arm provide greater benefit, compared with the component that is held constant across both the control and intervention arms. For example, some anxiety disorders are conceptualized as having several etiological components, including physiological arousal and a cognitive component in which threat appraisals or negative expectations contribute to symptoms [18]. Accordingly, RCTs may test a full package of cognitive-behavioral interventions and relaxation training against relaxation or cognitive restructuring only [19]. When the control and experimental conditions are comparable in terms of treatment format, implementation, and credibility, the research design permits strong causal inferences. However, such control conditions, which in their own right may have a substantial impact on the outcome, can require large sample sizes to be adequately powered. Often such studies are underpowered [20], leaving them vulnerable to concluding there is no difference across treatments when in fact there is (type II error).

Treatment component controls are potentially vulnerable to threats to internal validity that can overestimate the efficacy of a treatment. Particularly when testing a widely accepted intervention, clinician selection and allegiance effects can lead to inequality in therapist skill and enthusiasm across treatment arms. Lack of equipoise in treatment fidelity procedures may also aggravate these problems. Methods for preventing such biases include nesting clinicians within treatment arms, using clinician selection processes that ensure or evaluate allegiance to the different treatment arms, ensuring equivalence in the treatment fidelity monitoring across treatment arms, and monitoring clinician enthusiasm through, for example, outcome expectations [8].

Nonspecific Treatment Component Controls

Luminaries such as Meehl [21] and Paul [22] argued that RCTs for psychological interventions required some form of control condition analogous to the placebo controls used in medical research. Nonspecific control conditions commonly hold constant some or all of the following: (1) attention or amount of treatment contact, (2) human interaction variables, including clinician warmth and empathy, or social support and interaction among group participants, and/or (3) provision of a treatment rationale that controls for participant outcome expectations [23, 24].

As with specific treatment controls, nonspecific controls can have a substantial impact on many problems such as anxiety and mood disorders [19, 25], making it important to adequately power such trials. But nonspecific controls can also create vulnerabilities to threats to internal validity via treatment fidelity procedures. While control conditions for behavioral medicine interventions have often used relatively clearly defined, manualized programs [26], nonspecific controls in psychotherapy research have often not been manualized. When they are manualized, these conditions are often defined more by proscriptions (e.g. prohibitions against addressing change in cognitions or behaviors for a trial of a cognitive-behavioral intervention) than prescriptions for therapeutic procedures [19, 27]. The implicit assumption that nonspecific factors will emerge spontaneously is dubious, given the complexity of the requisite therapeutic skills.

Trials employing nonspecific control conditions often do not attain equipoise in monitoring across experimental and control conditions. Fidelity monitoring procedures typically focus on the prescribed features of the experimental treatment. Successful implementation of the control condition is often judged by the absence of specific factors rather than the presence of nonspecific factors [19]. If monitoring is used to provide feedback to the clinicians for the purposes of maintaining treatment fidelity, inequality in the level of prescription in treatment definition and monitoring may lead to inequality in clinician supervision and outcomes across treatment arms.

Finally, nonspecific control conditions are vulnerable to failures to maintain equipoise in clinician selection and allegiance. An investigator will likely have access to clinicians with expertise in and allegiance to the experimental treatment, but not to clinicians who identify themselves as providing 'nonspecific' treatment. Threats

Table 1. Advantages and disadvantages of control treatments

Comparison conditions	Uses, advantages	Possible disadvantages	Suggestions for implementation
Control conditions comprised of study-defined treatment			
Specific treatment component controls	• Most useful for understanding specific mechanisms • Control condition likely credible to participants • To the degree control treatment is a bona fide treatment, lower likelihood of ethical problems	• Vulnerability to failure of equipoise in treatment fidelity procedures • Vulnerability to failure of equipoise in clinician selection and allegiance • Specific treatment components may produce substantial effects, thereby requiring large sample sizes or increasing risk of type I error (rejecting valuable interventions) in RCTs	• Ensure treatments have equivalent levels of specification in manualization and prescriptive specificity, equivalent fidelity procedures, and equivalent feedback/supervision protocols • Equivalent rigor in clinician selection procedures, ensuring equivalent clinician allegiance, nesting clinicians in treatment arms, monitoring clinician allegiance and outcome expectations • Adequate sample sizes to detect potentially small, but meaningful effect sizes
Nonspecific treatment component controls	• Potentially useful for understanding mechanisms, testing effects of specific factors hypothesized to be beneficial	• The nonspecific factors present in the control treatment that are to be controlled for are often not clearly defined • Vulnerability to failure of equipoise in treatment fidelity procedures. Nonspecific interventions often proscribe specific clinician actions, but do not clearly prescribe clinician actions. Thus, nonspecific treatments are often ill-defined, fidelity is monitored only in terms of what is not performed, but not in terms of what is performed • Vulnerability to failure of equipoise in clinician selection and allegiance. Potential effects of decreased clinician outcome expectations in nonspecific control treatments • Nonspecific treatment component controls may produce substantial effects, thereby requiring large sample sizes or increasing risk of type I error (rejecting valuable interventions) in RCTs	• Nonspecific factors to be controlled for should be clearly identified based on the experimental treatment, theory, and the hypotheses • Equipoise in fidelity monitoring is achievable if treatment manuals have equal levels of prescriptive specificity, equivalent monitoring of prescribed clinician behaviors, and equivalent quality of clinician supervision • Ensure equivalence in clinician selection procedures and equivalent levels of clinician enthusiasm for the treatment. Assess clinician outcome expectations • Use adequate sample sizes
Control treatments provided outside study (e.g. TAU)	• To the degree TAU is representative, TAU may be a good way to evaluate if a treatment offers an improvement over current practice	• TAU can include many sources of variance (e.g. treatment, pt. identification, clinician knowledge and skills, follow-up procedures) that potentially threaten internal validity and diminish the ability to interpret results • To the degree TAU practices vary, it may be difficult to know if results generalize to a specific treatment setting • Vulnerability to failure of equipoise in treatment fidelity procedures. TAU conditions often do not include supervision of treatments • Vulnerability to failure of equipoise in clinician selection and allegiance. Experimental treatments often select clinicians, while TAU clinicians are not selected by the study • To the degree TAU is effective, it may require larger sample sizes	• Standardize procedures across treatment arms that might contribute to unwanted variance. Identify patients in TAU, educate TAU clinicians regarding care guidelines, schedule follow-up appointments, etc. Alternatively, monitor these factors and control for them statistically • Treatment provided in TAU should be monitored and described • Training, monitoring and supervision of TAU clinicians may be instituted. Alternatively, in effectiveness trials it may be preferable to minimize or eliminate some training and supervision procedures in the experimental treatment arm • Utilize clinicians in the treatment setting for both treatment arms
No treatment control, WLC	• More likely than many control conditions to produce strong effect sizes for experimental treatment • No-treatment controls less costly than other control conditions • Potentially useful for first evaluations of novel interventions	• Potential ethical problems if treatment is indicated for target problem • Vulnerability to control condition effects – may result in worsening of target problem • May cause patients to decline enrollment, thereby reducing feasibility and introducing sample bias • Different reasons for attrition across treatment arm resulting in different sample biases • Vulnerable to treatment fidelity procedure effects, against which little can be done	• Monitor health/well-being of participants • Assess and monitor potential threats to internal validity such as expectancies, help-seeking behavior and other services received • Continue assessment for patients dropping out of treatment

to clinician selection and allegiance may be managed by defining and implementing the control condition in a manner that suggests that it is a bona fide treatment, and by choosing clinicians whose levels of expertise and enthusiasm are equivalent to those of the therapists who will deliver the experimental treatment.

Control Conditions Using Treatments Provided as Part of Routine Clinical Care

A TAU control uses the routine intervention(s) ordinarily provided by clinicians in the settings from which participants are recruited. To the degree that the TAU condition is representative of current treatment practices, the RCT addresses the question, 'Would adding this treatment, or replacing TAU with this treatment, significantly improve outcomes in this setting?' This question is often of most interest to organizations that provide care.

While TAU controls for many traditional threats to internal validity, it has a number of potential problems resulting from the lack of investigator influence over the TAU intervention. First, it is often not clear what TAU is. The treatment provided by TAU clinicians may vary considerably across patients and providers [28, 29] and may not be adequately described. Since the effect size of an RCT equals '$ES_{experimental\ arm} - ES_{control\ arm}$', failure to characterize a TAU condition can make the results difficult to interpret and generalize (e.g. the treatment is effective compared to what?).

The outcomes of TAU may also include variability from sources other than the treatment itself. Consider the example of an RCT aimed at improving care for psychiatric disorders in primary care. TAU might be highly variable with respect to physician identification of the target disorder, knowledge of prescribing or referral guidelines, training and experience, or adequacy of follow-up [29, 30]. A significant finding may reflect these ancillary factors rather than the treatment itself. These unwanted sources of variance can be limited by standardizing them across treatment arms. For example, standardizing the identification of study participants across treatment arms can limit variance from identification procedures; providing standardized education of accepted treatment guidelines to all study providers can limit variability in provider knowledge; standardizing follow-up protocols can limit variability from differences in follow-up across treatment arms, and so on [31]. Such 'enhanced' TAU conditions can focus control on treatment effect. However, they also add elements to the clinical setting that change the way clinicians practice, thus reducing generalizability.

Another potential problem is that treatment arm may be confounded with fidelity monitoring (see Treatment Fidelity Procedures above). TAU conditions typically have no treatment manual, monitoring, or supervision of clinicians. If the experimental condition includes rigorous fidelity monitoring, it may be impossible to differentiate the effects of treatment from the effects of monitoring. Clinician selection and allegiance biases may also be introduced if clinicians in the two arms are not randomly assigned from the same pool of clinicians. These differences in clinician selection processes can reduce external validity and threaten internal validity when they result in differences across treatment arms on the level of therapist expertise, proficiency, allegiance, and/or enthusiasm. These potential weaknesses do not mean that TAU is not useful – most research designs have weaknesses – but they are factors to be considered in designing TAU conditions.

No-Treatment Control Conditions

The no-treatment control is a condition in which no alternative treatment is provided. The WLC condition is a close cousin, in which treatment is provided only after a period of time equivalent to or greater than the experimental treatment. No-treatment controls, used since the earliest psychotherapy trials [32], are intended to control for the traditional threats to internal validity and attempt to determine whether the experimental intervention is better than doing nothing.

Although no-treatment controls have an appealing simplicity, they also have a number of potential disadvantages. They may be ethically acceptable when the experimental treatment targets a problem without a treatment indication or when the trial focuses on a population with no immediate risks (e.g. prevention of depression) [33]. However, they may be less ethically acceptable when the trial targets severe disorders, for which effective treatment is both indicated and available.

Another potential problem with no-treatment control designs rests in the assumption that absence of treatment equates with absence of effect. As discussed above, control condition effects may threaten internal validity, as participants randomized to WLC conditions may improve less than would be expected compared to participants not enrolled in a trial [14, 15]. No-treatment controls are also particularly vulnerable to problems associated with treatment fidelity procedures effects and clinician selection and allegiance biases.

Of the control conditions discussed in this paper, no-treatment controls most often produce the largest effect

size for experimental treatments, because they are least likely to positively affect the outcome [34]. It might be argued that these control conditions set the bar too low, particularly given that many psychological targets will improve somewhat with almost any intervention [35]. Nevertheless, no-treatment conditions may be useful for the control of traditional threats to internal validity [6] and the detection of potential adverse effects of experimental interventions.

Other Control Conditions

The categories discussed above address the most commonly used forms of control conditions, but they are not exhaustive. For example, active control equivalence trials, comparing equivalence of two treatments, have been used in medical trials, but require very large sample sizes [36]. Pill placebo control conditions are often used when trials include a pharmacotherapy arm [37]. When clinical trialists of psychological interventions confront complex methodological problems, they also gain the opportunity to bring creativity and ingenuity to the design of control conditions, based on the study questions, feasibility issues, threats to internal validity, and other needs of the study.

Considerations in Selecting or Designing Control Conditions

As should be clear at this point, no RCT and no control condition is perfect. We will argue here that control condition selection or development should be influenced by the interaction of three factors: (1) considerations of statistical power and threats to internal validity, (2) RCT phase, and (3) the interests of stakeholders (e.g. patients, patient families, clinicians, payers, and researchers).

Statistical Power versus Control of Threats to Internal Validity

In an RCT, there is an implicit tradeoff between the statistical power to detect an effect and the level of control over threats to validity. In general, more comprehensive control conditions, such as nonspecific treatment component controls, control for more unwanted variability, focus effect sizes on the effects of the experimental treatment, and thereby reduce threats to internal validity. In contrast, control conditions that are less comprehensive in their control for threats to internal validity, such as no-treatment conditions, likely produce smaller within-group effects and larger between-group effects. This tradeoff between power and control of threats to internal validity is an inherent part of RCT design. Perfect control over threats to internal validity is not possible in the real world. Control over threats to validity is often sacrificed for other competing demands including feasibility, ethics, and statistical power.

RCT Phase

The experimental investigation of a psychological treatment evolves through a number of phases. The phases of RCTs have been described by the United States Food and Drug Administration and the National Institutes of Health [38] (see www.clinicaltrials.gov) for pharmaceutical trials. Similar phase models for the evaluation of psychological interventions have been proposed [39]. Phase I trials are feasibility trials aimed at treatment development, manual writing, exploring the potential effects of a new treatment, establishing parameters of the treatment such as number, length and frequency of sessions, safety testing [40, 41] and finalization of research protocols [42]. Phase II studies are preliminary trials conducted in a single setting with a clearly specified population. Phase III trials are large efficacy trials that typically include participants with comorbidities and multiple sites. Phase IV studies are effectiveness trials that evaluate the transportability of an intervention with demonstrated efficacy into the clinical setting and may evaluate additional questions such as optimal implementation systems or cost-effectiveness [43]. Sample sizes for phase I trials tend to be small, and increase as the investigation moves through the phases. Thus, power to detect effects also generally increases through the trial phases, although this may be offset to some degree by increased variability from expanded inclusion criteria and relaxation of rigid research protocols in phase IV RCTs.

Threats to Stakeholders

As the investigation of an intervention moves from phase I through phase IV trials, the sources of potential harm to stakeholders (e.g. patients, providers, payers, family members) shift. The potential harm in concluding that a treatment is not effective in any early phase of investigation when the treatment is in fact useful (type II error) could prevent or delay a beneficial treatment from receiving further investigation. On the other hand, the potential harm of supporting the validation of an ineffective treatment later in the process (type I error) could threaten public health and safety by allowing or encouraging an ineffective treatment to be accepted and implemented.

During the early phases of RCTs, the primary threat and potential harm to stakeholders is from a type II error. For example, failure to find support during phase II early efficacy trial for an intervention that is in fact effective could stop or delay clinical availability of a potentially useful intervention. The absence of a significant effect makes it difficult to publish findings, hard to obtain funding for further investigation, and may sap the enthusiasm of the investigative team to pursue further work. Thus, the importance of ensuring statistical power is high in these early phases. The importance of power, in light of threat of type II error to stakeholders, is only heightened by the fact that these early phase trials have smaller sample sizes than later phase trials.

In contrast, the potential harm from type I error during early phases of investigation is comparatively small, since broad implementation should not occur based on small preliminary studies. Requisite subsequent validation trials would provide additional opportunities to weed out ineffective treatments, which indeed happens frequently [44]. Thus, in early phase trials, statistical power is a primary concern. It may be both prudent and acceptable to sacrifice some level of control over threats to internal validity in favor of adequate power to detect any potential effects.

The threat or potential harm to stakeholders in later clinical trial phases – e.g. phases III and IV – shifts from type II to type I error. That is, the harm of finding a significant effect for an intervention when none in fact exists is substantial in later stages of research. A large, well-controlled trial is a substantial step towards acceptance by many stakeholders. As such, finding an intervention to be effective when it is in fact ineffective or harmful could promote the use of the treatment when better alternatives are available, resulting in wasted resources and/or harm to patients and consumers. Paradoxically, the larger sample sizes used at these later phases of testing increase power, thereby increasing the likelihood that statistical significance will be found, even for clinically meaningless effects [45]. Thus, in later clinical trial phases the importance of adequate control for threats to internal validity increases, while concerns regarding power generally decrease.

The Need for Guidelines

The foregoing discussion illustrates both the importance and complexities in the development and selection of control conditions, yet guidance for investigators is largely nonexistent. Psychological intervention research would benefit greatly from consensus statement on control conditions from a widely respected institution such as the NIH. We believe that in the absence of such a document, the following represent reasonable initial recommendations.

Decision-making regarding the selection or development of control conditions for RCTs of psychological interventions should systematically consider interactions between the RCT phase, statistical power and control of threats to internal validity in light of threats to stakeholders and the specific hypotheses of the RCT.

(1) Promote Innovation in Early Phase Trials

It is critical not to kill innovation in early trial phases. Phase I feasibility trials have small sample sizes that can be prone to biases, and accordingly should not be used for power or sample size calculations [42]. Issues of power deserve substantial weight in phase II early efficacy trials, even if this requires some sacrifices to internal validity. Accordingly, control conditions may control for some of the traditional threats to internal validity [6], but not necessarily all threats. For example, no-treatment (or WLC) control or a standard TAU that does not include enhancements may be acceptable, while treatment component controls may reduce power to the point of risking unwarranted findings of nonsignificance.

(2) Strengthen Controls for Threats to Internal Validity in Later Phase Trials

As interventions move to phase III and IV trials, protecting stakeholders from ineffective treatments becomes paramount. These trials also typically employ larger sample sizes that can accommodate rigorous controls for threats to internal validity, including threats not usually considered. Threats to internal validity from treatment fidelity procedures can be minimized by ensuring that experimental and control arms use treatment manuals with similar levels of prescriptive specificity, and that interventionists receive comparable training, monitoring, and supervision, or that treatment fidelity monitoring is minimized across all arms. To minimize clinician selection and allegiance effects, clinicians should be selected with the same level of care and, if possible, allegiance across treatment arms. When possible, clinicians should be blinded to whether they are delivering the experimental or the control treatment. Negative effects of a control condition on outcomes should at least be measured and controlled for statistically. When these threats to internal validity can-

not be controlled for, this should be clearly articulated in any report.

(3) Use of Nonspecific Controls Should Not Be Automatic, but Should Be Guided by the Hypothesis

Whether or not nonspecific factors should be controlled has been a topic of heated debate for decades [4, 5]. Most investigators would likely agree that nonspecific control conditions are not necessary or appropriate for phase I or II trials, as this would constitute a nearly insurmountable hurdle in moving novel treatments from an initial development into more rigorous testing. The use of nonspecific controls in later phase trials depends principally on the specific hypotheses, and how the nonspecific factors are conceptualized in the treatment and hypotheses. If nonspecific factors are conceptualized as an integral part of the treatment and the aim of the trial is to test the efficacy of the treatment, it is not advisable to organize the control arm around nonspecific factors. However, to the degree that nonspecific factors are not conceptualized as part of the treatment, or that the hypotheses focus on specific mechanisms, it may be useful to control for nonspecific factors. For example, in behavioral medicine, exercise has long been believed to reduce fatigue and boost energy. A trial aimed at testing the efficacy of an exercise promotion intervention in reducing fatigue would likely not require control for nonspecific treatment components, particularly if the question were simply whether or not the treatment works. However, investigators have hypothesized that interventions aimed at reducing fatigue and increasing energy through exercise are effective through specific biological pathways. If these pathways are part of the hypotheses, nonspecific factors associated with treatment delivery such as attention or social interaction in group interventions would be considered incidental factors to be controlled for [46]. Under these circumstances, where hypotheses include specific mechanisms that exclude nonspecific factors, the use of nonspecific controls may be both appropriate and necessary. In short, the use of nonspecific treatment control conditions should support and be linked to the aims and hypotheses of the trial.

(4) When Using a Nonspecific Component Control, Minimize Effects from Treatment Fidelity Procedures

A number of measures can be employed to prevent threats to internal validity arising from nonspecific component controls: (1) clearly identify the nonspecific factors in the experimental treatment that are to be the focus of the control condition, (2) provide clear prescriptive instructions for clinicians about how to implement nonspecific factors, in addition to any prohibitions, (3) implement fidelity monitoring equivalently across treatment arms for prescribed and proscribed clinician behaviors, and (4) strive for equivalence in therapist selection procedures and allegiance.

(5) Content of TAU Conditions Should Be Clear

The content of TAU conditions can vary considerably across providers, sites, and studies. TAU procedures should at least be monitored and clearly described.

(6) Consider Potential Threats to Internal Validity a priori

Different disorders, problems, and treatments may be more or less vulnerable to different types of threats to internal validity. For example, different disorders may vary in the degree to which their symptoms change over time, or the likelihood that they may drop out of different types of treatments or control conditions. As an illustration of the latter, the number of participants who drop out of no-treatment, WLC, or weak TAU control conditions is similar, or even slightly lower, than dropout from experimental treatments among patients with depression in depression and anxiety disorders [27, 30]. On the other hand, dropout from WLC or TAU control conditions tends to be greater than in experimental treatments among RCTs involving schizophrenics [47] and substance users [48]. Such information is critical in considering control condition options. Development of empirical evidence on the relevance of factors affecting internal validity would inform decisions regarding control condition design and selection.

(7) Conflicts of Interest

While not specifically the focus of this paper, the influence of investigator allegiance has been well documented [11]. RCTs should include plans to manage such conflicts of interest, for example by including co-investigators of different allegiances in planning and monitoring.

Conclusion

In the universe of effect sizes that make up our RCT evidence base for psychological interventions, control conditions remain dark matter, exerting effects that are unseen, ill-defined, and for the most part unquantified. While there have been disputes over which control condi-

tions should be used, such debates have produced more heat than light.

This discussion of control groups may appear abstract to the practicing clinician. But the effects of control conditions permeate every corner of clinical applications of psychological interventions. Inappropriate control conditions can overestimate the effectiveness of a treatment, or kill off a potentially useful treatment. Thus, the question of control conditions can have far-ranging implications, from writing grant proposals to publishing RCT results, from what is taught in clinical graduate programs to what is practiced by clinicians using evidence-based practices, and from what is demanded by patients to what is supported by third-party payers and promoted by policy makers. Good control conditions can provide solid evidence upon which to build practice and policy. It is therefore critical that we begin to develop principles and guidelines that can promote appropriate use and design of control conditions in the evaluation of psychological interventions. Such efforts should consider all threats to internal validity, balance promotion of innovation with the need to protect the public from ineffective or dangerous treatments, and provide a framework that supports and standardizes investigators' process of making decisions and choices regarding control conditions.

References

1 Chambless DL, Ollendick TH: Empirically supported psychological interventions: controversies and evidence. Annu Rev Psychol 2001;52:685–716.
2 Cuijpers P, van Straten A, Warmerdam L, Smits N: Characteristics of effective psychological treatments of depression: a meta-regression analysis. Psychother Res 2008;18:225–236.
3 Kazdin AE: Methodology, design, and evaluation in psychotherapy research; in Bergin AE, Garfield SL (eds): Handbook of Psychotherapy and Behavior Change, ed 4. New York, Wiley, 1994.
4 Kirsch I: Placebo psychotherapy: synonym or oxymoron? J Clin Psychol 2005;61:791–803.
5 Lohr JM, Olatunji BO, Parker L, DeMaio C: Experimental analysis of specific treatment factors: efficacy and practice implications. J Clin Psychol 2005;61:819–834.
6 Campbell DT, Stanley JC: Experimental and Quasi-Experimental Designs for Research. Chicago, Rand McNally, 1966.
7 Shadish WR, Cook TD, Campbell DT: Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Boston, Houghton Mifflin, 2002.
8 Bellg AJ, Borrelli B, Resnick B, Hecht J, Minicucci DS, Ory M, Ogedegbe G, Orwig D, Ernst D, Czajkowski S: Enhancing treatment fidelity in health behavior change studies: best practices and recommendations from the NIH Behavior Change Consortium. Health Psychol 2004;23:443–451.
9 Crits-Christoph P, Baranacki K, Kurcias JS, Beck AT, Carroll K, Perry K, Luborsky L, McLellan T, Woody GE, Thompson L, Gallagher D, Zitrin C: Meta-analysis of therapist effects in psychotherapy outcome studies. Psychother Res 1991;1:81–91.
10 Cottraux J, Note I, Yao SN, de Mey-Guillard C, Bonasse F, Djamoussian D, Mollard E, Note B, Chen Y: Randomized controlled comparison of cognitive behavior therapy with Rogerian supportive therapy in chronic post-traumatic stress disorder: a 2-year follow-up. Psychother Psychosom 2008;77:101–110.
11 Luborsky L, Diguer L, Seligman DA, Rosenthal R, Krause ED, Johnson S, Halperin G, Bishop M, Berman JS, Schweizer E: The researcher's own therapy allegiances: a 'wild card' in comparisons of treatment efficacy. Clin Psychol Sci Pract 1999;6:95–106.
12 Gaudiano BA, Herbert JD: Methodological issues in clinical trials of antidepressant medications: perspectives from psychotherapy outcome research. Psychother Psychosom 2005;74:17–25.
13 Walach H, Bosch H, Lewith G, Naumann J, Schwarzer B, Falk S, Kohls N, Haraldsson E, Wiesendanger H, Nordmann A, Tomasson H, Prescott P, Bucher HC: Effectiveness of distant healing for patients with chronic fatigue syndrome: a randomised controlled partially blinded trial (EUHEALS). Psychother Psychosom 2008;77:158–166.
14 Harris KB, Miller WR: Behavioral self-control training for problem drinkers: components of efficacy. Psychol of Addict Behav 1990;4:82–90.
15 Schmidt MM, Miller WR: Amount of therapist contact and outcome in a multidimensional depression treatment program. Acta Psychiatr Scand 1983;67:319–332.
16 Basham RB: Scientific and practical advantages of comparative design in psychotherapy outcome research. J Consult Clin Psychol 1986;54:88–94.
17 Hardy GE, Barkham M, Shapiro DA, Reynolds S, Rees A, Stiles WB: Credibility and outcome of cognitive-behavioural and psychodynamic-interpersonal psychotherapy. Br J Clin Psychol 1995;34:555–569.
18 Barlow DH: Anxiety and Its Disorders: The Nature and Treatment of Anxiety and Panic. New York, Builford, 2002.
19 Borkovec TD, Costello E: Efficacy of applied relaxation and cognitive-behavioral therapy in the treatment of generalized anxiety disorder. J Consult Clin Psychol 1993;61:611–619.
20 Foa EB, Dancu CV, Hembree EA, Jaycox LH, Meadows EA, Street GP: A comparison of exposure therapy, stress inoculation training, and their combination for reducing posttraumatic stress disorder in female assault victims. J Consult Clin Psychol 1999;67:194–200.
21 Meehl PE: Psychotherapy. Annu Rev Psychol 1955;6:129–131.
22 Paul GL: Behavior modification research: design and tactics; in Franks CM (ed): Behavior Therapy: Appraisal and Status. New York, McGraw-Hill, 1966, pp 29–62.
23 Kazdin A: Research Design in Clinical Psychology. New York, Harper & Row, 1980.
24 Frank JD, Frank JB: Persuasion and Healing: A Comparative Study of Psychotherapy, ed 3. Baltimore, Hopkins University Press, 1991.
25 Mohr DC, Hart SL, Julian L, Catledge C, Honos-Webb L, Vella L, Tasch ET: Telephone-administered psychotherapy for depression. Arch Gen Psychiatry 2005;62:1007–1014.
26 Jacob RG, Shapiro AP, O'Hara P, Portser S, Kruger A, Gatsonis C, Ding Y: Relaxation therapy for hypertension: setting-specific effects. Psychosom Med 1992;54:87–101.
27 Ward E, King M, Lloyd M, Bower P, Sibbald B, Farrelly S, Gabbay M, Tarrier N, Addington-Hall J: Randomised controlled trial of non-directive counselling, cognitive-behaviour therapy, and usual general practitioner care for patients with depression. 1. Clinical effectiveness. BMJ 2000;321:1383–1388.

28 Unutzer J, Katon W, Callahan CM, Williams JW Jr, Hunkeler E, Harpole L, Hoffing M, Della Penna RD, Noel PH, Lin EH, Arean PA, Hegel MT, Tang L, Belin TR, Oishi S, Langston C: Collaborative care management of late-life depression in the primary care setting: a randomized controlled trial. JAMA 2002;288:2836–2845.
29 Mohr DC, Likosky W, Bertagnolli A, Goodkin DE, Van Der Wende J, Dwyer P, Dick LP: Telephone-administered cognitive-behavioral therapy for the treatment of depressive symptoms in multiple sclerosis. J Consult Clin Psychol 2000;68:356–361.
30 Simon GE, Ludman EJ, Tutty S, Operskalski B, Von Korff M: Telephone psychotherapy and telephone care management for primary care patients starting antidepressant treatment: a randomized controlled trial. JAMA 2004;292:935–942.
31 Reynolds CF 3rd, Degenholtz H, Parker LS, Schulberg HC, Mulsant BH, Post E, Rollman B: Treatment as usual (tau) control practices in the prospect study: managing the interaction and tension between research design and ethics. Int J Geriatr Psychiatry 2001;16:602–608.
32 Gelder MG, Marks IM: Severe agoraphobia: a controlled prospective trial of behaviour therapy. Br J Psychiatry 1966;112:309–319.
33 Peden AR, Rayens MK, Hall LA, Beebe LH: Preventing depression in high-risk college women: a report of an 18-month follow-up. J Am Coll Health 2001;49:299–306.
34 Atkins CJ, Kaplan RM, Timms RM, Reinsch S, Lofback K: Behavioral exercise programs in the management of chronic obstructive pulmonary disease. J Consult Clin Psychol 1984;52:591–603.
35 Herbert JD, Gaudiano BA: Moving from empirically supported treatment lists to practice guidelines in psychotherapy: the role of the placebo concept. J Clin Psychol 2005;61:893–908.
36 Wellek S: Testing Statistical Hypotheses of Equivalence. Boca Raton, CRC Publishers, 2002.
37 DeRubeis RJ, Hollon SD, Amsterdam JD, Shelton RC, Young PR, Salomon RM, O'Reardon JP, Lovett ML, Gladis MM, Brown LL, Gallop R: Cognitive therapy vs. medications in the treatment of moderate to severe depression. Arch Gen Psychiatry 2005;62:409–416.
38 US Food and Drug Administration: Title 21, section 321.21, phases of an investigation: Code of Federal Regulations. 2005, 21CFR312.21.
39 Rounsaville BJ, Carroll KM, Onken LS: A stage model of behavioral therapies research: getting started and moving on from stage I. Clin Psychol Sci Pract 2001;8:133–142.
40 Lilienfeld SO: Psychological treatments that cause harm. Perspect Psychol Sci 2007;2:53–70.
41 Mohr DC: Negative outcome in psychotherapy: a critical review. Clin Psychol Sci Pract 1995;2:1–27.
42 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.
43 Glasgow RE, Davidson KW, Dobkin PL, Ockene J, Spring B: Practical behavioral trials to advance evidence-based behavioral medicine. Ann Behav Med 2006;31:5–13.
44 Velazquez EJ, Califf RM: All that glitters is not gold. Lancet 2000;355:1568–1569.
45 Kazdin AE: The meanings and measurement of clinical significance. J Consult Clin Psychol 1999;67:332–339.
46 Dishman RK, Berthoud HR, Booth FW, Cotman CW, Edgerton VR, Fleshner MR, Gandevia SC, Gomez-Pinilla F, Greenwood BN, Hillman CH, Kramer AF, Levin BE, Moran TH, Russo-Neustadt AA, Salamone JD, Van Hoomissen JD, Wade CE, York DA, Zigmond MJ: Neurobiology of exercise. Scand J Med Sci Sports 2006;16:470.
47 Bola JR, Mosher LR: Treatment of acute psychosis without neuroleptics: two-year outcomes from the Soteria project. J Nerv Ment Dis 2003;191:219–229.
48 Spooner C, Mattick RP, Noffs W: Outcomes of a comprehensive treatment program for adolescents with a substance-use disorder. J Subst Abuse Treat 2001;20:205–213.