Practical Behavioral Trials to Advance Evidence-Based Behavioral Medicine

Russell E. Glasgow, Ph.D.
Kaiser Permanente Colorado

Karina W. Davidson, Ph.D.
Columbia College of Physicians and Surgeons

Patricia L. Dobkin, Ph.D.
McGill University

Judith Ockene, Ph.D.
University of Massachusetts Medical School

Bonnie Spring, Ph.D.
University of Illinois Chicago and
Hines Hospital, VAMC
Hines, Illinois

ABSTRACT

There is a well-documented gap between research and practice in many areas of behavioral medicine. This gap is due in part to limitations in the capacity of the research database to address questions that are of central concern to clinicians, administrative decision makers, and policymakers. Thus, there has been a call for "practical clinical trials" that compare clinically viable alternative interventions and assess multiple outcomes important for clinical and policy decisions in diverse patient populations and settings. Such trials offer great potential, and they raise interesting challenges regarding optimal research design and source of funding. We discuss issues related to practical clinical trials in behavioral medicine, propose a need for practical behavioral trials (PBTs), and describe design features that will facilitate clinical and policy decision making. This type of PBT can help to close the gap between research and practice and advance the field of evidence-based behavioral medicine. We discuss potential challenges and objections to PBTs and conclude by providing recommendations for the design, conduct, reporting, and review of practical trials.

(**Ann Behav Med** 2006, *31*(1):5–13)

INTRODUCTION

There is an increasingly well-documented gap between best practices identified in research and the care delivered in practice for both preventive services (1) and chronic illness (2). There are multiple reasons for this gap, but at least part of the cause is that many practitioners do not view the majority of research studies as being applicable to their clinical situations. Discussions of evidence-based behavioral medicine (EBBM) are recent (3,4), but it seems important to ensure that the EBBM movement considers ways in which the evidence base can help to close the gap between research and practice.

There has been considerable debate in the field of EBBM about the bases for rating research quality and whether randomized controlled efficacy trials always represent the best form of evidence (5–8) or are inapplicable to some clinical and community situations (9). Such debates, although interesting, have generally produced more heat than light. We on the Society of Behavioral Medicine EBBM Committee suggest this is not an either/or issue. Instead, a more useful contribution can be made by understanding the strengths and limitations of the information that can be obtained from a range of designs. Thorough testing and evaluation of behavioral interventions requires the use of many research approaches, designs, and questions. That is, no single design can adequately capture the totality of information needed to assess the usefulness of any given intervention. Both efficacy and effectiveness designs (10,11) contribute to the continuum of research required. Interventions first should be tested for optimal intensity, timing, mode of action (or mechanism), tolerability, and safety (Phase 1 trial). The next step is testing for preliminary efficacy—that an intervention has the expected impact on the primary outcome(s), when controlling for other possible determinants of the effect (Phase 2 trial). Next, larger trials to determine the efficacy of an intervention against placebo, credible active alternatives, or attention controls, are performed (Phase 3a trial).

In this article, we describe the next phase of trials (Phase 3b) needed to address the external validity, policy implications, and usefulness of the intervention. These trials continue to be controlled, usually for community/standard alternative interventions, so that the question of usefulness and external validity can be properly addressed. All phases are needed. When inter-

---

This article is a product of the Evidence-Based Behavioral Committee of the Society of Behavioral Medicine (SBM; see http://www.sbm.org/ebbm/index.html). Appreciation is expressed to Drs. Michael Goldstein and Evelyn Whitlock and to Kimberlee Trudeau, M.A. for their helpful comments on an earlier version of this article.

Reprint Address: R. E. Glasgow, Ph.D., Kaiser Permanente Colorado, Clinical Research Unit, 335 Road Runner Lane, Penrose, CO 81240. E-mail: russg@ris.net

ventions have known mechanisms, expected effect sizes, and regularly outperform reasonable controls, we still need to determine if the intervention will be effective in practice—something that is seldom done (12). We call for the development of methods and criteria that can make these last types of randomized clinical trials more practical and relevant for clinicians and policymakers.

In an important article on "practical clinical trials" (PCTs), Tunis, Stryer, and Clancy (13) argued cogently that an increasing willingness of health care clinicians and policy decision makers to base decisions on scientific evidence is being stymied by a lack of research data that address the questions decision makers need to have answered. They proposed a number of key characteristics of these trials that can increase their relevance. This article extends the ideas and recommendations of Tunis and colleagues (13) to behavioral medicine and discusses recommendations for "practical behavioral trials" (PBTs) appropriate for behavioral medicine that can accelerate the transfer of research into practice (11). By a *behavioral* trial, we mean a study in which either (a) the intervention employs behavioral strategies, procedures, or theory, or (b) the primary outcomes involve behavior change on the part of patients, clinicians, families, or larger systems (e.g., change in worksite policies).

As summarized in Table 1, there are five key characteristics of PCTs (13) and a suggested set of three additional characteristics of PBTs. Like earlier discussions of efficacy versus effectiveness trials (10,11) and of "pragmatic" trials in medicine (14–17), seldom is a trial a pure efficacy trial or a pure PBT. Rather, we view the extent to which a trial has PBT characteristics as a continuum and hope that future research *that is intended to address clinical and public health issues* will employ more of these features than is currently the case (12).

Although having many similarities to dissemination (e.g., Phase 4) studies (18,19), PCTs and PBTs are probably best considered a subset of Phase 3 trials, because they include randomization and a control condition. Unlike some effectiveness studies, PBTs are intended to address critical issues in adoption or decision making, study multiple outcomes, and compare clinically meaningful alternatives, and they do not necessarily need to be extremely large or expensive studies.

TABLE 1
Key Characteristics of Practical Behavioral and Clinical Trials

1. Answer questions of key stakeholders (e.g., clinicians, decision makers, and policymakers).
2. Assess multiple and relevant outcomes including cost, generalization, and quality of life—see Table 2.
3. Compare clinically meaningful alternatives—using research designs matched to state of knowledge.
4. Recruit a diverse, heterogeneous sample and evaluate robustness across key subgroups.
5. Include multiple, representative settings and interventionists.
6. Issues especially important in practical behavioral trials:
   a. Training: Specify level of training/expertise necessary and amount of training provided.
   b. Address patient preferences.
   c. Provide algorithms for intervention tailoring.

KEY ISSUES IN PRACTICAL CLINICAL TRIALS

Research That Answers Questions Posed by Key Stakeholders

Despite great progress that has helped translate basic science discoveries into new treatments and technologies, there is a striking dearth of research that helps stakeholders decide which procedures warrant their investment (13). Stakeholders whose decision making is impeded by this lack of evidence exist at all levels in the health care system. For any ailment, patients must decide which of a number of alternative treatments offers them the best balance among: likely benefit on a valued outcome, low or tolerable risk, and affordable cost. Clinicians need to determine for any particular patient which treatment offers the best risk–benefit ratio. Usually they need to make such choices in the absence of adequate knowledge about long-term effectiveness, comparative efficacies of alternative treatments, and whether most patients will find the treatment regimen sufficiently appealing to adhere and accrue the intended benefits. In addition, clinicians need to gauge the likelihood that they will be reimbursed for providing any particular service.

The next tier of those impeded by lack of evidence includes systematic reviewers. Ideally, their work should form the basis for practice guidelines that endorse implementation and reimbursement of some treatments more strongly than others. How should reviewers make recommendations for standards of care when the evidence base has research design flaws and lacks information about representative populations or outcomes like cost-effectiveness, harms, functional status, and quality of life? Finally, payers need to decide, in a context of limited resources, whether a new, more expensive treatment yields sufficient advantages compared with less expensive alternatives to warrant the added investment. How are payers to make such decisions on the basis of studies that compared each treatment only to placebo, no care, or usual care, rather than directly comparing alternative treatments to each other?

Measurement and Analyses for PBTs

The key characteristics of measures for practical trials are that they are multiple and address issues of importance to clinicians, decision makers, and policymakers. We propose that a package of measures of behavioral change, quality of life, implementation, generalization, and economic outcomes should be included in PBTs (see Table 2). Such a package is feasible to include in many PBTs. The last three types of measures can be included without adding burden to participant assessments. Failure to adequately implement an intervention is one of the most frequent reasons that interventions do not work in real-world settings and should be documented to facilitate interpretation of results (20,21). Evaluations should include multiple intervention staff to assess robustness and should assess adaptations or "re-invention" that occurs over time (22,23) to make a protocol more practical. The purpose of including heteroge-

TABLE 2
Recommended Measures for Practical Behavioral Trials

1. Behavior change: at multiple levels if relevant (e.g., patient, clinician, system).
2. Quality of life or potential unintended consequences.
3. Implementation of intervention components by representative clinical staff.
4. Generalization across participants, settings, intervention staff, and outcomes.
5. Cost and economic outcomes.

neous settings and patients in PBTs is to evaluate the generalizability of the intervention (or alternatively, to assess "moderator") effects of settings, patient, and interventionist characteristics (5,24). Such analyses greatly aid decision makers in deciding whether an intervention is appropriate to their setting. The costs of delivering the practical intervention in representative settings should be documented, and it is preferable that other economic outcomes, such as cost-effectiveness analyses, also be included. Questions about cost and related economic issues are among the first questions decision makers raise about successful behavioral interventions. Yet seldom do we have answers to these questions (12). Although comprehensive economic analyses can themselves be expensive and perhaps beyond the scope of some studies, most PBTs should, at minimum, be able to present information about costs required to deliver or replicate an intervention (25).

The first two categories of measures in Table 2, behavioral outcomes and quality of life, do require additional time from participants but are the central outcomes most relevant to the majority of PBTs. Behavior change should be assessed using practical but state-of-the-art methods (26). If the PBT is evaluating an intervention to change health care provider or staff behavior or produce system change, then it is critical to measure not only patient behavior but also that of the staff or system (e.g., policy implementation). Quality of life and harms are proposed as a final set of outcomes appropriate for PBTs. Collection of such data both helps to capture unanticipated negative outcomes and provides a common metric for comparing the impact of interventions for different conditions (27,28). The RE-AIM model (29,30) offers one approach for combining the five categories of measures proposed earlier in an integrated fashion (http://www.re-aim.org). This planning and evaluation framework organizes these issues into the dimensions of *R*each (participation), *E*ffectiveness, *A*doption (by representative settings and intervention staff), *I*mplementation, and *M*aintenance.

Compare Clinically Meaningful Treatment Alternatives Using Research Designs Adequate to Compare Them

The guiding principle in choosing comparison arms in practical trials is to provide key stakeholders with data they need to decide *among clinically viable alternatives*. In medicine there is frequently a current standard of care against which a new treatment can be compared. Too often, a new generation of trials supporting a new treatment fails to perform comparisons against these already-established treatments. The consequence is that key stakeholders in the reimbursement of medical practice are left without crucial information about the relative cost-effectiveness, benefit, and risk profiles of competing possible treatments. In contrast, for behavioral medicine, often there is currently no viable standard or competing treatment for many behavioral problems. As emphasized previously, earlier trials that answer questions about efficacy over placebo, attention, and other kinds of control conditions should already have been conducted. Similarly, dismantling studies, in which the active ingredient(s) of the behavioral intervention is determined, also have a role in the progression leading to a PBT. For example, before a PBT is undertaken, it may be important to determine whether stress management can be effective with 4 sessions rather than 16 sessions, or whether the intervention can be delivered as effectively over the telephone as in person. Then a PBT can be conducted on an intervention that is of the minimum intensity (and cost) needed to produce change in the average patient seen in a typical practice setting (31). The question to be answered by a PBT is, should this intervention be adopted? To answer this question, the control condition should be the currently available usual or standard care for that condition. In cases where the current standard of care is actually no treatment, that is the standard against which appropriate outcome and economic analyses can be benchmarked.

For some behavioral interventions, it is unclear whether some level of current intervention constitutes the usual standard of care. For example, physical activity promotion interventions, such as a brief physician counseling or school setting redesign, currently are applied inconsistently. In such a case the investigative team needs to decide whether these practices are sufficiently common to serve as a comparison condition or whether no treatment is the comparison that stake holders will find more meaningful.

Finally, there are some areas, such as smoking cessation, for which a clear standard of behavioral care exists. Here, there exist practice guidelines directing providers to offer brief counseling that implements the 5 As: ask if the patient smokes, advise quitting, assess readiness to make a quit attempt, assist in supplying access to appropriate resources, and arrange follow-up (32). Moreover, when it comes to providing assistance to quit, guidelines also offer clear recommendations regarding which treatments, both behavioral and pharmacological, constitute first- and second-line standards of care. Because often such recommendations are not implemented in practice, another decision for practical trial design often comes down to whether a newly proposed behavioral treatment can be compared to a single comparison condition (prescribed best practices) or whether a second control arm (usual care) is also required.

"Noninferiority" randomized controlled trial (RCT) designs (33) are appropriate for situations in which the prospect of not offering a viable treatment to all participants raises ethical concerns. However, a recent decision by the Food and Drug Administration (34) highlights some of the dangers presented by a

noninferiority design. By employing two active treatment arms, such a design calculates power to determine that the new treatment is "not inferior" to the standard treatment. In theory, such a design has much to offer. It appears on the face ethical, because patients are not assigned to receive placebo or no care. However, the assumptions on which a design rests are at times tenuous and the analyses complex. One presumption is that the standard of care has been well tested and consistently been found better than placebo or other control conditions. Although this point seems obvious, it is not necessarily the case in practice. An old adage in psychopharmacology advises that doctors should "use the new drugs quickly, while they still work." Because past trials were conducted in a different time, with different patients, different settings, and perhaps even somewhat different forms of the same disorders, it cannot necessarily be assumed that contemporary RCTs would yield the same evidence of efficacy as was previously attained.

For reasons such as these, a three-arm trial including some variant of placebo control may be prudent even when there is a well-accepted standard of care, unless it can be firmly established that the current treatment standard is robust against historical threats and confounds and demonstrably superior to contemporary placebo or no treatment controls.

Recruit a Diverse, Heterogeneous Sample of Participants

It is important that a practical trial recruit diverse groups of participants and not exclude the more complicated, multiple-problem, underserved, lower educational attainment or lower health literacy (35) patients who are of concern to primary-care clinicians and policymakers. Increasing the representativeness of study participants requires using fewer exclusion criteria. Broadening inclusion and narrowing exclusion criteria enables studies to enroll patients who represent the range and distribution of those seen in clinical practice, including higher risk and multimorbid patients, who may respond differently than lower risk patients to certain treatments. Efficient methods to obtain more heterogeneous samples without the expense associated with random sampling include sampling from a defined population, stratified sampling, or purposive sampling (5). A particular issue receiving increased attention currently is medical patients who are also depressed. In the past, such patients have often been explicitly or implicitly (e.g., through the use of "run-in" adherence periods) screened out of intervention studies.

Conduct the Study in Multiple, Representative Settings

In the past, most controlled studies were conducted in a small number of tertiary-care, university, or specialty settings. Understandably, practitioners asked questions about the extent to which one can generalize from such studies. They raised questions regarding whether the interventionists, staff expertise, resources, workloads, and time dedicated to a given trial versus other duties were comparable enough to allow generalization to more typical practice settings (36). To strengthen the evidence base of EBBM, studies need to be conducted in the multiple and diverse settings that comprise the clinical practice of behavioral medicine. Medicine has recently addressed this issue at least partially through the formation of Practice-Based Research Networks that consist of collaborations of practicing clinicians in primary-care community settings who are interested in participating in research (32,37). Formation of similar networks of behavioral medicine settings might function in much the same way.

ISSUES SPECIFIC TO PBTs

In addition to the characteristics shared by all PCTs (see Tables 1 and 2), several other factors influence the practicality of a trial in behavioral medicine. First, the level of professional accreditation and training that an interventionist requires to successfully deliver a treatment should already have been established in previous trials. Some behavioral interventions require no training (e.g., tailored e-mails, printouts, or telephone responses generated by an expert computer system). Others (e.g., problem-solving treatment or motivational interviewing) require specific training in the therapy to be delivered as well as a prior skill base often equivalent to bachelor's- or master's-level training in a human services profession. Still other more complex interventions (e.g., dialectical behavioral therapy for persons with borderline personality disorder) require advanced clinical training and credentialing. Establishing these levels, communicating them, and setting criteria that operationalize adequacy of training are key to the successful testing, costing, and then translation of PBTs.

Standardization and dissemination of the training required to administer the behavioral intervention is crucial. At the high end of training intensity are programs that involve 1 or more days of training plus continued, ongoing quality control. At the lower end of intensity is simple Web posting of the required material. Consideration of the necessary level of training, reporting on the amount and cost of the training, and access to a successful training program are other key elements that render a trial practical and its results more generalizable.

Patient preferences have been of particular interest to those studying behavioral interventions (38,39) and are of central relevance to PBTs. Some data suggest that U.S. adults may prefer to receive behavioral rather than medical treatments for problems such as depression (40,41). If a patient has a strong bias for or against medication use, or is against discussing early family-of-origin issues, such preferences should not only be considered but also ideally woven into the choice of PBT design (42). An example of how it is possible to incorporate patient preferences is provided by the IMPACT study, a large trial of depression treatment that involved 18 primary-care sites (43). Patients were randomized to receive either usual physician care for depression/dysthymia or their choice of antidepressant medication or psychotherapy. This type of added design element may meaningfully augment the way in which behavioral interventions can be not just tested but sustained after the trial results are known and can answer practical questions about options that represen-

tative patients are most likely to choose. Other research designs have been developed to assess differences in outcomes when patients are allowed to choose an intervention versus being randomly assigned to it (42).

A third key element for PBTs is intervention tailoring. PBT protocols need to offer clinicians a way to augment, step up, or adapt interventions in light of how the patient is progressing. The use of tailoring algorithms offers the potential to individualize patient treatment while making explicit the treatment principles being followed in a trial. In essence, the algorithm spells out how clinical decisions should be made at key treatment points in a manner that integrates data about patient progress. An example has been provided by Loisel and colleagues (44–46), who conducted a population-based RCT on back pain management in the industries surrounding the city of Sherbrooke, Quebec, Canada (a small city whose universal health care system also serves the rural population within its catchment area). This trial was designed to prevent prolonged disability in injured workers with subacute back pain. Workers (N =130) from eligible workplaces in the Sherbrooke area (n = 31) who had been absent from work for more than 4 weeks for back pain were randomized, based on their workplace (i.e., a cluster randomization design), to one of four treatment arms: (a) usual medical care; (b) clinical rehabilitation intervention (clinical examination by a back pain medical specialist, participation in a back school *after 8 weeks* of absence from regular work, and, *if necessary*, a multidisciplinary work rehabilitation intervention—with a psychologist or occupational therapist who oversaw a progressive return to regular tasks *after 12 weeks absence from work*); (c) occupational intervention (occupational physician visits and a "participatory ergonomics evaluation" in the workplace, which involved an ergonomist, the patient, his or her supervisor, and union/employer representatives to determine need for job modifications); and (d) full intervention, that is, an integrated combination of (b) and (c) (the "Sherbrooke Model").

Pertinent to PBTs, the clinical intervention included, after 8 weeks absence from work, a visit to a back pain specialist and back school. Moreover, after 12 weeks' absence, treatment was stepped up to include a *multidisciplinary* intervention (i.e., fitness development), "work hardening," and cognitive–behavioral treatment. The results demonstrated that the integrated clinical–occupational model was effective in increasing the rate of return to regular work more than twofold, compared to usual medical care. The study also illustrates that PBTs need not be huge or inordinately expensive. Most important, a 6-year follow-up study showed that the fully integrated disability prevention model saved more sick days on benefits than the usual care or partial interventions. The mean costs (1998 Canadian dollars) during the follow-up period (M = 6.4 years)—that is, the combined costs of the intervention and consequences of disease costs—were (a) $33,079 (intervention = $9,562; consequences = $23,517), (b) $16,902 (intervention = $6,857; consequences = $10,045), (c) $16,252 (intervention = $3,432, consequences = $12, 820), and (d) $14,494 (intervention = $7,434, consequences = $7,060) and supported the Sherbrooke model.

EXAMPLES OF PBTs

From the preceding descriptions, it might appear impossible or overwhelming to conduct a PBT that addresses the variety of issues recommended. However, the Sherbrooke study we just described, and the two examples that follow, illustrate that PBTs can be of moderate size and quite feasible.

An example of a PBT was conducted by Ockene and colleagues (47,48), who tested a smoking treatment program for low-income pregnant women in community health centers. This study, "Quit Together," was designed to evaluate the effect of a provider counseling and office system intervention in obstetric, pediatric, and special supplemental nutrition programs for women, infants, and children (WIC) clinics on smoking and relapse rates among pregnant and postpartum women. The following sections summarize how the Quit Together trial addressed several key PCT characteristics (13).

Compare New Treatments to Realistic Alternative Interventions

There was no evidence that primary-care practitioners in community health centers serving low-income pregnant women were performing the recommended standard of care, that is, brief counseling that implements the 5 As (49). Given the lack of information about the delivery of the gold standard, this trial did not compare new treatments to guideline-based intervention but investigated the delivery of the recommended standard of care compared to usual care. The intervention consisted of three components: (a) provider training to deliver a smoking intervention based on national clinical practice guidelines (49) tailored to the woman's stage of change and delivered through three channels (obstetric, pediatric, and WIC providers); (b) an office practice management system to routinely screen for smoking status, prompt/remind providers to intervene, document the encounter, distribute materials, and arrange follow-up; and (c) establishment of program boards to coordinate the transfer of documentation among clinics, including periodic meetings with representatives from all clinics. Regular clinic staff were expected to remind clinicians to deliver the intervention. There were no paid research assistants.

Include Multiple Outcomes Relevant to Key Stakeholders

Important measures in this trial were the extent to which each group of clinicians (i.e., obstetricians, pediatricians, and nutritionists) performed the 5 As. These measures were important to understand how the intervention was implemented and the amount of intervention received. Also measured were smoking cessation, reduction of tobacco use, stages of change, and maintenance of cessation by women who had quit at the time they learned of their pregnancy.

Recruit a Diverse, Heterogeneous Sample of Participants

The trial was designed to recruit current smokers or those who had quit with pregnancy and were from low-income minor-

ity populations. The intervention group was about 25% White, non-Hispanic; 39% Black, non-Hispanic; and 28% Hispanic, thus representing the diverse population of low-income women seen in community health centers. In addition, 47% had less than a high school education.

Conduct the Study in Multiple Representative Settings

There were six community health centers randomized to intervention or usual care. As is often a challenge in PBTs, it was difficult to match the centers, because the population of each center and the system used to run the centers were each different.

DIABETES QUALITY IMPROVEMENT PBT

Glasgow and colleagues (50) conducted a PBT to evaluate a computer-assisted diabetes management intervention to help patients and their health care teams to increase the quality of patient-centered care. The practical question addressed was how typical fee-for-service primary-care practices could efficiently improve the quality of diabetes care they provided. The study involved a diverse sample of 886 type 2 diabetes patients who were generally *representative* of diabetes *patients* in Colorado. The study was conducted in the offices of 52 mixed-payer primary-care physicians throughout the state, illustrating study of *multiple settings*. The intervention was conducted via touch screen computer and *by regular office staff*, who were trained by research staff in onsite training. These *heterogeneous* staff members consistently delivered the various intervention components, including physician advice, goal setting, and follow-up phone calls. In this study, the *practical alternative* consisted of "enhanced usual care," which involved a touch screen computer-assisted health risk appraisal and feedback to patient and provider but not the other treatment components. The program was significantly better than the randomized comparison condition on *multiple measures*. For example, it was significantly better in improving completion of both laboratory checks for which patients were due, and it improved receipt of patient-centered counseling for self-management goals. Measures of depression and diabetes-related *quality of life* showed improvements in both conditions, but the interventions did not differ significantly on those outcomes.

Barriers to PBTs

As Tunis et al. (13) suggested, a primary reason for the dearth of PCTs, and PBTs in particular, is the shortage of public and private sector funding for them. Because they can require large sample sizes and extended follow-up, PCTs can be expensive (as can traditional efficacy and effectiveness trials). Tunis et al. cited as one example the National Emphysema Treatment Trial, which enrolled 1,200 patients and cost at least $35 million. Yet the National Institutes of Health lack a specific mechanism to identify high-priority questions that health care decision makers need to have resolved and that might warrant PCTs/PBTs. Addressing such decision-making needs is the mission of the Agency for Healthcare Research and Quality, which devotes approximately $30 million annually to PCTs (13) but lacks additional resources to expand that effort. With $55 million allocated annually to PCTs, the Veterans Affairs Cooperative Studies Program has carried out some important PCTs on, for example, coronary artery bypass graft surgery and arthroscopic knee surgery for osteoarthritis. Note, though, that none of these large-scale PCTs has involved a behavioral treatment condition. To the extent that PCTs aim to help clinical decision makers determine the best procedures to maximize health outcomes, minimize risks, and manage costs, we urge that this state of affairs be changed. That realization is receiving attention in countries with universal health care. For example, as discussed earlier, in Canada, a PBT implementing multidisciplinary treatment within community-based settings for injured workers (the majority having low back pain) has been supported by the Quebec government (44). A population-based study is currently underway to determine whether the prevention model is effective for construction workers.

Some of the logistical and structural issues that need to be addressed to conduct PBTs involve very practical considerations. To maximize generalizability, it would be best for in-house clinic staff to provide the intervention, but that will be challenging unless their usual workload is lightened or alternatives such as interactive technology, are used. There also may be other bureaucratic obstacles, such as union rules that hinder staff ability to work after hours (e.g., conducting family therapy during the evenings) or to carry out functions outside their usual duties (e.g., registered nurses training inpatients to perform progressive relaxation prior to surgery). Another question is whether research data can be integrated into medical charts so as to emulate clinical practice and whether that can be accomplished while meeting human participant and Health Insurance Portability Assurance Act requirements. Alternatively, if an outside team administers the intervention or collects the data, the research may interfere with usual clinic practice and, therefore, be unlikely to be maintained.

These challenges can be addressed with careful planning. Inclusion of clinical colleagues in target settings from the very beginning conceptualization of the PBT using principles of community-based participatory research (51,52) is also important. These participatory or action research principles emphasize active involvement, from the outset of planning, of clinical partners and the decision makers who need to act on the results. Input from staff in the settings where the PBT is to be conducted is critical to its success, not only because they can sabotage it if they are against it but also because they know their clientele and the logistics of working within their setting. Also, their buy-in will be essential if the intervention is to be sustained and become part of routine clinical practice.

RECOMMENDATIONS

PBTs require cooperation among scientists, practitioners, policy advisors, funding agencies, professional organizations, journal editors, community and clinical interventions sites, and of course patients. Although such a consensus is never easy and requires careful attention to an inclusionary process, all parties will benefit from having the type of PBT research data that can move behavioral medicine interventions clearly into the arena of an evi-

dence-based practice. There is an urgent need to disseminate effective behavioral medicine programs into practice, and we view PBTs as an important avenue to accomplish this priority. As L. W. Green said, "If we want more evidence-based practice, then we need more practice-based evidence" (53, p. 15). We also need to present results in terms that use less jargon and are in the language of clinicians, decision makers, and policymakers. Specific actions that would support the conduct of more PBTs and enhance adoption and dissemination might be formation of (a) networks of behavioral practitioners engaged primarily in clinical practice and interested in practical research, similar to the practice-based primary care research networks (37), (b) an SBM special interest group on PBTs, and (c) coordinated programs of PBTs (supported by studies leading up to a PBT) rather than expecting a single study to answer all PBT questions regarding a given program.

A challenge encountered when reviewing behavioral medicine literature is that different investigative groups measure different outcomes and do so with measures that are differentially sensitive to change. Therefore, as discussed earlier and elsewhere (36), we recommend greater use of standardized measures in the five domains presented in Table 2. We especially recommend collection of data on quality of life, functional status, and harms that can be directly compared across diverse interventions and target populations (27,28). We also recommend that behavioral trials more consistently include measures of cost and economic outcomes. Health care resources are limited, and responsible selection among alternative uses of limited funds requires knowledge of the costs as well as benefits of alternative interventions (27,28,54,55). We realize that trials are conducted for multiple purposes and that a given study may not be able to include all the PBT elements recommended. However, many PBT elements can be included without major increases in expense or time needed to conduct a trial. Increasing attention to these issues will greatly facilitate adoption of EBBM programs.

Focusing attention on the importance of issues central to the conduct and reporting of PBTs cannot be separated from review issues. Recommendations for researchers to report on the factors discussed earlier will be unsuccessful if journal and grant reviewers are not informed of, and do not consider, these issues in their evaluations (12,22,56). We will see increases in the number and quality of PBTs only when study sections and journals evaluating EBBM trials start to include such criteria and provide feedback on the extent to which such factors are integrated into grant proposals and research reports.

CONCLUSIONS

The actions advocated in this article would result in changes in the way that we conduct at least some portion of behavioral trials. We are not advocating a restriction on other kinds of research. Rather, we advocate balanced consideration of the kinds of issues that PBTs and PCTs address to achieve greater gains toward translating research into practice. We are optimistic that our field will become a leader in this type of practical, evidence-based health care research, in much the same way that the groundbreaking methodological approaches to the design of smoking cessation intervention trials spawned more sophisticated and informed subsequent intervention research (57). To accomplish the SBM's mission to "promote the health of individuals, families, communities, and populations" and to have evidence relevant to the health care policy and clinical decisions facing our society, more PBT research is needed.

REFERENCES

(1) McGlynn EA, Asch SM, Adams J, et al.: The quality of health care delivered to adults in the United States. *New England Journal of Medicine.* 2003, *348:*2635–2645.

(2) Institute of Medicine, Committee on Quality of Health Care in America: *Crossing the Quality Chasm: A New Health System For the 21st Century.* Washington, DC: National Academy Press, 2001.

(3) Davidson KW, Trudeau KJ, Ockene JK, Orleans CT, Kaplan RM: A primer on current evidence-based review systems and their implications for behavioral medicine. *Annals of Behavioral Medicine.* 2004, *28:*226–238.

(4) Spring B, Pagoto S, Kaufmann PG, et al.: Invitation to a dialogue between researchers and clinicians about evidence-based behavioral medicine. *Annals of Behavioral Medicine.* 2005, *30:*125–137.

(5) Shadish WR, Cook TD, Campbell DT: *Experimental and Quasi-Experimental Design for Generalized Causal Inference.* Boston: Houghton Mifflin, 2002.

(6) Stirman SW, DeRubeis RJ, Crits-Christoph P, Rothman A: Can the randomized controlled trial literature generalize to nonrandomized patients? *Journal of Consulting and Clinical Psychology.* 2005, *73:*127–135.

(7) Blair E: Gold is not always good enough: The shortcomings of randomization when evaluating interventions in small heterogeneous samples. *Journal of Clinical Epidemiology.* 2004, *57:*1219–1222.

(8) Brouwers MC, Johnston ME, Charette ML, et al.: Evaluating the role of quality assessment of primary studies in systematic review of cancer practice guidelines. *BMC Medical Research Methodology.* 2005, *5:*8.

(9) Parker G: Evaluating treatments for the mood disorders: Time for the evidence to get real. *Australia New Zealand Journal of Psychiatry.* 2004, *38:*408–414.

(10) Flay BR: Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. *Preventive Medicine.* 1986, *15:*451–474.

(11) Glasgow RE, Lichtenstein E, Marcus AC: Why don't we see more translation of health promotion research to practice? Rethinking the efficacy to effectiveness transition. *American Journal of Public Health.* 2003, *93:*1261–1267.

(12) Glasgow RE, Klesges LM, Dzewaltowski DA, Bull SS, Estabrooks P: The future of health behavior change research: What is needed to improve translation of research into health promotion practice? *Annals of Behavioral Medicine.* 2004, *27:*3–12.

(13) Tunis SR, Stryer DB, Clancy CM: Practical clinical trials: Increasing the value of clinical research for decision making in clinical and health policy. *Journal of the American Medical Association.* 2003, *290:*1624–1632.

(14) Coyne JC, Palmer SC, Shapiro PJ: Prescribing antidepressants to advanced cancer patients with mild depressive symptoms is not justified. *Journal of Clinical Oncology.* 2004, *22:*205–206.

(15) Fisch MJ, Loehrer PJ, Kirsteller J, et al.: Fluoxetine versus placebo in advanced cancer outpatients: A double-blinded trial of

the Hoosier Oncology Group. *Journal of Clinical Oncology.* 2003, *21:*1937–1943.
(16) Schwartz D, Lellouch J: Explanatory and pragmatic attitudes in therapeutical trials. *Journal of Chronic Disease.* 1967, *20:*637.
(17) Vallve C: A critical review of the pragmatic clinical trial. *Medicina Clinica.* 2003, *121:*384–388.
(18) Secker-Walker RH, Holland RR, Lloyd CM, Pelkey D, Flynn BS: Cost effectiveness of a community based research project to help women quit smoking. *Tobacco Control.* 2005, *14:*37–42.
(19) Epping-Jordan JE: Research to practice: International dissemination of evidence-based behavioral medicine. *Annals of Behavioral Medicine.* 2004, *28:*81–87.
(20) Bellg AJ, Borrelli B, Resnick B, et al.: Enhancing treatment fidelity in health behavior change studies: Best practices and recommendations from the Behavior Change Consortium. *Health Psychology.* 2004, *23:*443–451.
(21) Basch CE, Sliepcevich EM, Gold RS: Avoiding Type III errors in health education program evaluations. *Health Education Quarterly.* 1985, *12:*315–331.
(22) Rotheram-Borus MJ, Flannery ND: Interventions that are CURRES: Cost-effective, useful, realistic, robust, evolving, and sustainable. In Rehmschmidt H, Belfer ML, Goodyer I (eds), *Facilitating Pathways: Care, Treatment, and Prevention in Child and Adolescent Health.* New York: Springer, 2004, 235–244.
(23) Rogers EM: *Diffusion of Innovations* (5th Ed.). New York: Free Press, 2003.
(24) Cronbach LH, Glesser GC, Nanda H, Rajaratnam N: *The Dependability of Behavioral Measurements: Theory of Generalizability for Scores and Profiles.* New York: Wiley, 1972.
(25) Meenan RT, Stevens VJ, Hornbrook MC, et al.: Cost-effectiveness of a hospital-based smoking cessation intervention. *Medical Care.* 1998, *36:*670–678.
(26) Glasgow RE, Ory MG, Klesges LM, et al.: Practical and relevant measures of health behavior for primary care settings. *Annals of Family Medicine.* 2005, *3:*73–81.
(27) Kaplan RM: Quality of life, resource allocation, and the U.S. health-care crisis. In Dimsdale JE, Baum A (eds), *Quality of Life in Behavioral Medicine Research.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc., 1995, 3–30.
(28) Kaplan RM: The significance of quality of life in health care. *Quality of Life Research.* 2003, *12:*3–16.
(29) Glasgow RE: Evaluation of theory-based interventions: The RE-AIM model. In Glanz K, Lewis FM, Rimer BK (eds), *Health Behavior and Health Education* (3rd Ed.). San Francisco: Wiley, 2002, 531–544.
(30) Klesges LM, Estabrooks PA, Glasgow RE, Dzewaltowski D: Beginning with the application in mind: Designing and planning health behavior change interventions to enhance dissemination. *Annals of Behavioral Medicine.* 2005, *29:*66–75.
(31) Croyle, R. [Personal communication]. 2003.
(32) Agency for Healthcare Research and Quality: *Evidence-Based Practice.* Retrieved December 20, 2005, from http://www.ahcpr.gov/clinic/epcix.htm.2001
(33) D'Agostino Sr. RB, Massaro JM, Sullivan IM: Non-inferiority trials: Design concepts and issues—The encounters of academic consultants in statistics. *Statistics in Medicine.* 2003, *22:*169–186.
(34) Is that it, then, for blockbuster drugs? [Editorial]. *The Lancet.* 2004, *364:*1100.
(35) Schillinger D, Grumbach K, Piette JD, et al.: Association of health literacy with diabetes outcomes. *Journal of the American Medical Association.* 2002, *288:*475–482.
(36) Glasgow RE, Magid DJ, Beck A, Ritzwoller D, Estabrooks PA: Practical clinical trials for translating research to practice: Design and measurement recommendations. *Medical Care.* 2005, *43:*551–557.
(37) Nutting PA, Beasley JW, Werner JJ: Practice-based research networks answer primary care questions. *Journal of the American Medical Association.* 1999, *281:*686–689.
(38) King M, Nazareth I, Lampe F, et al.: Impact of participant and physician intervention preferences on randomized trials: A systematic review. *Journal of the American Medical Association.* 2005, *293:*1089–1099.
(39) Rokke PD, Scogin F: Depression treatment preferences in younger and older adults. *Journal of Clinical Geropsychology.* 1995, *1:*243–257.
(40) Chilvers C, Dewey M, Fielding K, et al.: Antidepressant drugs and generic counselling for treatment of major depression in primary care: Randomised trial with patient preference arms. *British Medical Journal.* 2001, *322:*772–775.
(41) Guthrie E: Psychotherapy for patients with complex disorders and chronic symptoms. *British Journal of Psychiatry.* 2000, *177:*131–137.
(42) Bradley C: Clinical trials—Time for a paradigm shift? *Diabetic Medicine.* 1988, *5:*107–109.
(43) Williams Jr. JW, Katon W, Lin EH, et al.: The effectiveness of depression care management on diabetes-related outcomes in older patients. *Annals of Internal Medicine.* 2004, *140:*1015–1024.
(44) Loisel P, Durand MJ, Diallo B, et al.: From evidence to community practice in work rehabilitation: The Quebec experience. *Clinical Journal of Pain.* 2003, *19:*105–113.
(45) Loisel P, Abenhaim L, Durand P, et al.: A population-based, randomized clinical trial on back pain management. *Spine.* 1997, *22:*2911–2918.
(46) Loisel P, Lemarie J, Poitras S, et al.: Cost–benefit and cost-effectiveness analysis of a disability prevention model for back pain management. *Occupational and Environmental Medicine.* 2002, *59:*807–815.
(47) Ockene JK, Ma Y, Zapka JG, et al.: Spontaneous cessation of smoking and alcohol use among low-income pregnant women. *American Journal of Preventive Medicine.* 2002, *23:*150–159.
(48) Pbert L, Ockene JK, Zapka J, et al.: A community health center smoking-cessation intervention for pregnant and postpartum women. *American Journal of Preventive Medicine.* 2004, *26:*377–385.
(49) Fiore M, Bailey W, Cohen S, et al.: *Smoking Cessation: Clinical Practice.* Guideline No. 18, AHCPR No. 96-0692. Rockville, MD: U.S. Department of Public Health and Human Services, Public Health Service, Agency for Health Care Policy and Research, 1996.
(50) Glasgow RE, Nutting PA, King DK, et al.: A practical randomized trial to improve diabetes care. *Journal of General Internal Medicine.* 2004, *19:*1167–1174.
(51) Green LW: From research to "best practices" in other settings and populations. *American Journal of Health Behavior.* 2001, *25:*165–178.
(52) Leung MW, Yen IH, Minkler M: Community-based participatory research: A promising approach for increasing epidemiology's relevance in the 21st century. *International Journal of Epidemiology.* 2004, *33:*499–506.

(53) Green LW, Ottoson JM: *From efficacy to effectiveness to community and back: Evidence-based practice vs. practice-based medicine*. "From Clinical Trials to Community: The Science of Translating Diabetes and Obesity Research." Washington, DC: 2004.
(54) Gold MR, Siegel JE, Russell LB, Weinstein MC: *Cost-Effectiveness in Health and Medicine*. New York: Oxford University Press, 2003.
(55) Lamm RD: *The Brave New World of Health Care* (1st Ed.) Golden, CO: Fulcrum, 2004.
(56) Des Jarlais DC, Lyles C, Crepaz N, TREND Group: Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: The TREND statement. *American Journal of Public Health.* 2004, *94:*361–366.
(57) Best A, Hiatt RA, Cameron R, Rimer BK, Abrams DB: The evolution of cancer control research: An international perspective from Canada and the United States. *Cancer Epidemiology, Biomarkers and Prevention.* 2003, *12:*705–712.