

Why Don't We See More Translation of Health Promotion Research to Practice? Rethinking the Efficacy-to-Effectiveness Transition

Russell E. Glasgow, PhD, Edward Lichtenstein, PhD, and Alfred C. Marcus, PhD

The gap between research and practice is well documented. We address one of the underlying reasons for this gap: the assumption that effectiveness research naturally and logically follows from successful efficacy research. These 2 research traditions have evolved different methods and values; consequently, there are inherent differences between the characteristics of a successful efficacy intervention versus those of an effectiveness one. Moderating factors that limit robustness across settings, populations, and intervention staff need to be addressed in efficacy studies, as well as in effectiveness trials. Greater attention needs to be paid to documenting intervention reach, adoption, implementation, and maintenance. Recommendations are offered to help close the gap between efficacy and effectiveness research and to guide evaluation and possible adoption of new programs. (*Am J Public Health*. 2003;93:1261–1267)

Despite a growing literature documenting prevention and health promotion interventions that have proven successful in well-controlled research, few of these interventions are consistently implemented in applied settings. This is true across preventive counseling services for numerous target behaviors, including tobacco use, dietary change, physical activity, and behavioral health issues (e.g., alcohol use, depression). Several recent reviews and meta-analyses have documented this gap,^{1,2} and the task forces on both clinical preventive services and community preventive services have noted that in several areas there is insufficient applied evidence available to make recommendations at present.^{3–5} Most of the Healthy People 2000 objectives⁶ were not met, and the even more ambitious goals in Healthy People 2010 are similarly unlikely to be met without significant changes in the status quo.^{7,8} To meet these challenges, we will need to have substantially more demonstrations of how to effectively implement recommendations in typical settings and in locations serving minority, low-income, and rural populations facing health disparities.

This situation is not unique to preventive interventions, as strikingly documented in the recent Institute of Medicine report *Crossing the Chasm*,⁹ which summarizes the similar state of affairs regarding many medical and disease management interventions. For example, there

is increasing consensus on evidence-based diabetes management practices to prevent complications and on the importance and cost-effectiveness of these practices.¹⁰ However, these recommendations—and especially those related to lifestyle counseling and behavioral issues—are poorly implemented in practice.^{11–14}

This gap between research and practice is the result of several interacting factors, including limited time and resources of practitioners, insufficient training,¹⁵ lack of feedback and incentives for use of evidence-based practices, and inadequate infrastructure and systems organization to support translation.^{8,16} In this article, we focus on another reason for the slow and incomplete translation of research findings into practice: the logic and assumptions behind the design of efficacy and effectiveness research trials.

EFFICACY AND EFFECTIVENESS TRIALS

Many of the methods used in current prevention science are based on 2 influential papers published in the 1980s: Greenwald and Cullen's¹⁷ description of the phases of cancer control research and Flay's analysis of efficacy and effectiveness research.¹⁸ Both papers argued for a logical progression of research designs through which promising intervention

ideas should proceed. These papers had many positive effects in helping to establish prevention research and enhancing acceptability among other disciplines. However, they may also have had an important and inadvertent negative consequence that derives from the assumption that the best candidates for effectiveness studies—and later dissemination—are interventions that prove successful in certain types of efficacy research. We argue that this assumption, or at least the way in which it has been operationalized over the past 15 years, has often led to interventions that have low probability of success in real-world settings.

To understand this point, it is necessary first to briefly review the seminal papers by Flay¹⁸ and Greenwald and Cullen.¹⁷ Efficacy trials are defined by Flay as a test of whether a “program does more good than harm when delivered under optimum conditions.”^{18(p451)} Efficacy trials are characterized by strong control in that a standardized program is delivered in a uniform fashion to a specific, often narrowly defined, homogeneous target audience. Owing to the strict standardization of efficacy trials, any positive (or negative) effect can be directly attributed to the intervention being studied.

Effectiveness trials are defined as a test of whether a “program does more good than harm when delivered under real-world conditions.”^{18(p451)} They typically standardize availability and access among a defined population while allowing implementation and levels of participation to vary on the basis of real-world conditions. The primary goal of an effectiveness trial is to determine whether an intervention works among a broadly defined population. Effectiveness trials that result in no change may be the result of a lack of proper implementation or weak acceptance or adherence by participants.^{18,19}

Greenwald and Cullen¹⁷ proposed 5 phases of intervention research presumed to unfold in

a sequential fashion. This continuum begins with Phase I research to formulate and develop intervention hypotheses for future study. Phase II studies develop methodologies that can be used in future efficacy or effectiveness studies. Phase III (efficacy) studies test intervention hypotheses, using methods that have been tested in Phase II. Thus, Phase III studies are designed to test interventions for efficacy, with an emphasis on internal validity, the purpose of which is to establish a causal link between the intervention and outcomes. Given this emphasis on internal control, Greenwald and Cullen note that Phase III studies can be conducted in settings and with samples that will “optimize interpretation of efficacy,” including study samples that may be more homogeneous than the ultimate target population, and settings that will maximize management of and control over the research process.

The main objective of Phase IV (effectiveness) studies is to measure the impact of an intervention when it is tested within a population that is representative of the intended target audience. Given that Phase IV studies should yield results that are generalizable, there is also the presumption that the context and setting for delivering the intervention should likewise be generalizable to the intended program users. In Phase V studies, effective Phase IV interventions are translated into large-scale demonstration projects. The major concern is implementation fidelity of an intervention that will now be introduced within even broader populations, including entire communities. This final phase (dissemination research), where collaboration and coordination with various community partners is likely to receive even greater attention, is intended to provide the necessary data and experience to move interventions into public health service programs at the national, regional, state, and local levels.

Greenwald and Cullen specifically advocated that intervention research unfold in a systematic fashion, building on and extending the body of science accumulated in previous phases. By explicitly defining the difference between Phase III and Phase IV research as being an emphasis on internal control versus representativeness, both Flay and Greenwald and Cullen assumed that successful Phase III trials would lead naturally to Phase IV trials. Unfortunately, this has not occurred.^{1,11,20} Instead, we

currently find ourselves in a situation in which we have many small-scale efficacy studies of unknown generalizability and few successful effectiveness trials.^{21,22} In particular, we know very little about the representativeness of participants, settings, or intervention agents participating in health promotion research.^{1,21}

Although the National Cancer Institute no longer emphasizes this linear “phases of research” model,^{23,24} the model was extremely influential in guiding an entire generation of research; many researchers, reviewers, and editors still use this framework when designing, funding, and evaluating research—and in deciding what types of studies are needed to advance a given area. Similar phase models are influential in evaluating prevention effectiveness²⁵ and in developing drug therapies. In the remainder of this article, we discuss how this well-intentioned and logical phase of research paradigm may have fallen short of its intended goal, and propose approaches to remedy the present situation.

Our primary thesis is that this “trickle-down” model of how to translate research into practice—namely, that the optimal way to develop disseminable interventions is to progress from efficacy studies to effectiveness trials to dissemination projects—is inherently flawed, or at least incomplete. We posit that given the respective cultures, values, and methodological traditions that have developed within efficacy versus population-based effectiveness research, it is highly unlikely

that interventions that are successful in efficacy studies will do well in effectiveness studies, or in real-world applications.

Table 1 summarizes the key characteristics of well-designed efficacy and effectiveness trials, using the RE-AIM evaluation framework.^{26,27} This model for evaluating interventions is intended to refocus priorities on public health issues, and it gives balanced emphasis to internal and external validity (see <http://www.re-aim.org>). RE-AIM is an acronym for Reach, Efficacy or Effectiveness (depending on the stage of research), Adoption, Implementation, and Maintenance.

Reach refers to the participation rate among those approached and the representativeness of participants. Factors determining reach are the size and characteristics of the potential audience and the barriers to participation (e.g., cost, social and environmental context, necessary referrals, transportation, and inconvenience). *Efficacy or effectiveness* pertains to the impact of an intervention on specified outcome criteria and includes measures of potential negative outcomes as well as intended results (as recommended by Flay,¹⁸ but seldom collected)^{28,29} (D.A. Dziewaltowski et al., unpublished data, 2002). *Adoption* operates at the setting level and concerns the percentage and representativeness of organizations or settings that will conduct a given program. Rogers³⁰ has written extensively on adoption and dissemination issues. Factors associated with adoption include political and cultural fit,

TABLE 1—Distinctive Characteristics of Efficacy and Effectiveness Intervention Studies, Using RE-AIM^{26,27} Dimensions for Program Evaluation

RE-AIM Issue	Efficacy Studies	Effectiveness Studies
Reach	Homogeneous, highly motivated sample; exclude those with complications, other comorbid problems	Broad, heterogeneous, representative sample; often use a defined population
Efficacy or effectiveness	Intensive, specialized interventions that attempt to maximize effect size; very standardized; randomized designs	Brief, feasible interventions not requiring great expertise; adaptable to setting; randomized, time series, or quasi-experimental designs
Adoption	Usually 1 setting to reduce variability; settings with many resources and expert staff	Appeal to and work in multiple settings; able to be adapted to fit setting
Implementation	Implemented by research staff closely following specific protocol	Implemented by variety of different staff with competing demands, using adapted protocol
Maintenance and cost	Few or no issues; focus on individual level.	Major issues; setting-level maintenance is as important as individual-level maintenance

cost, level of resources and expertise required, and how similar a proposed service is to current practices of an organization. *Implementation* refers to intervention integrity, or the quality and consistency of delivery. Finally, *maintenance* operates at both the individual and the setting or organizational level. At the individual level, maintenance refers to how well behavior changes hold up in the long term. At the setting level, it refers to the extent to which a treatment or practice becomes institutionalized in an organization.

Table 1 summarizes how the RE-AIM dimensions apply to the efficacy–effectiveness distinction. Efficacy trials typically limit reach by seeking motivated, homogeneous participants with minimal or no complications or comorbidities. The considerable degree of initial screening for eligibility inherently limits the reach of an efficacy trial. Adoption is often treated as a nonissue for efficacy trials so long as at least one or, in some trials, a few settings are willing to participate. For effectiveness trials, reach is usually higher because participants are drawn from a broad and “defined” population. Adoption is critical because the settings need to commit their own resources and expect the intervention to “fit” with existing procedures.

Implementation in an efficacy trial is usually accomplished by research staff following a standardized protocol, whereas in an effectiveness trial, regular staff with many competing demands on their time must implement the intervention. While such staff are also guided by a protocol, adherence is likely to be more variable.¹ Because they are implemented by research staff, efficacy interventions are often more complex and intensive than effectiveness interventions. Maintenance is usually a nonissue for efficacy trials at the setting level; it is expected that the intervention will cease when final assessments are completed and research staff depart. Since effectiveness trials are intended to represent typical setting conditions, it is hoped that the intervention will be maintained, assuming there are positive results.

WHY THE DISCONNECT?

We conclude that the characteristics that cause an intervention to be successful in efficacy research (e.g., intensive, complex, highly

standardized) are fundamentally different from, and often at odds with, programs that succeed in population-based effectiveness settings (e.g., having broad appeal, being adaptable for both participants and intervention agents). If this is the case, then the “system” of moving from research to usual service programs, to which we have subscribed, may be broken and may need to be substantially modified.

Why does this linear progression of research, which is analogous to the steps used successfully to evaluate and bring pharmaceuticals to market, seem to fail with behavioral and health promotion research? One contextual factor is that, before trials, pharmaceutical companies invest considerable time and money establishing that the drug affects relevant biological mediators to a much greater extent than behavioral researchers invest in showing that their interventions affect psychosocial mediators. Granted, industry has vastly more resources. But we suggest that key differences also reside in the nature of the interventions.

Standard medical interventions (e.g., drugs or surgery) are presumed to be robust, readily transferable from setting to setting, and to work approximately equally across broad categories of patients. Clinicians exercise discretion about dosage and surgeons vary in experience, but it is still presumed that the pill is the same whoever administers it. Medicinal and surgical protocols can be relatively precisely defined, and adherence to them can be more easily monitored relative to behavioral interventions. Behavioral interventions are more difficult to define and standardize in part because of the inherent interactivity with client characteristics, preferences, and behaviors. This is exacerbated when behavioral interventions are delivered by staff whose training and expertise fall outside of behavioral science. In efficacy trials, research staff usually bring expertise in behavioral intervention and ensure that it is implemented consistently. This level of quality control and standardization is typically absent among regular health care staff implementing interventions for effectiveness trials.

There are 2 underlying differences between efficacy and effectiveness approaches that we feel are responsible for the current state of affairs. The first is that in an effort to enhance

internal validity and control extraneous factors, the tradition in efficacy studies has been to simplify and narrow settings, conditions, participants, and a variety of other factors. There is nothing inherently wrong with this methodological approach, and the tradition of reductionism (e.g., understanding effects by isolating them and removing or controlling other factors) has contributed much to the advancement of science and theory.³¹ The problem is that usually the longer-range intent is to generalize beyond the narrow conditions of the efficacy trial. In effectiveness trials, an intervention must be robust across a variety of different participants, settings, conditions, and other less controlled factors. Equally important, it must appeal to a broad “defined population” or target audience.

A classic example of the typical differences between a health care efficacy study and an effectiveness trial concerns subject selection. In a tightly controlled efficacy trial, only highly motivated, homogenous self-selected volunteers who do not have any complications or other comorbid conditions are eligible (to control for potential confounding factors). Then, following success in such an efficacy study, we expect the same intervention to appeal to and be effective in a much broader cross-section of participants, many of whom have comorbid conditions and may not volunteer for treatment.

The second key difference between efficacy and effectiveness trials concerns how settings and contextual factors are treated. In efficacy studies, the usual approach is to control variance by restricting the setting to one set of circumstances—for example, one particular clinic (which often includes intervention experts). In contrast, a key characteristic of effectiveness trials is to produce robust effects and to understand variation in outcomes across heterogeneous settings and delivery agents. Therefore, it should not be surprising when the results of an intervention are efficacious under a highly specific set of circumstances but fail to replicate across a wide variety of settings, conditions, and intervention agents in effectiveness research.

SHALL THE TWAIN EVER MEET?

From the above discussion, it may seem hopeless to expect congruence across findings

from efficacy and effectiveness studies. Some might go so far as to suggest, as one reviewer of this manuscript did, that perhaps efficacy studies should be abandoned altogether. We are optimistic, however, that there are solutions to the present disconnect. In brief, we need to embrace and study the complexity of the world, rather than attempting to ignore or reduce it by studying only isolated (and often unrepresentative) situations.³² What is needed is a “science of larger social units”³³ that takes into account and analyzes the social context(s) in which experiments are conducted. To advance our present state of science, the question that we need to ask of both efficacy and effectiveness studies is “What are the characteristics of interventions that can (a) reach large numbers of people, especially those who can most benefit, (b) be broadly adopted by different settings (worksite, school, health, or community), (c) be consistently implemented by different staff members with moderate levels of training and expertise, and (d) produce replicable and long-lasting effects (and minimal negative impacts) at a reasonable cost?”

This suggested focus has important implications. It implies that we need to consider not only individual participants but also the settings within which they reside and receive treatment. This move to a multilevel approach is consistent with developments in several fields, and methodologies for how to handle such factors are available. There is not only a rich conceptual history to the study of generalization³⁴ and of representative or purposeful sampling,^{35,36} but also statistical methods for handling these contextual factors.³⁷

This comes down to an issue of generalization.³⁸ The prevailing view seems to be that efficacy studies should focus only on internal validity and theoretical process mechanisms, and that issues of external validity should be left until later effectiveness studies. In contrast, we argue that issues of moderating variables (external validity) need to be addressed in *both* efficacy and effectiveness studies. Brewer³⁹ conceptualizes such social context factors as moderating variables that influence the conclusions that can be drawn about the efficacy of an intervention. Moderating variables (e.g., race/ethnicity, socioeconomic status, type of setting or intervention agent) are

relatively stable factors that interact with the intervention or change the effect of the program. Researchers should consider elevating hypotheses related to moderator variables to primary aims.

WHAT CAN BE DONE? DISCUSSION AND RECOMMENDATIONS

It is difficult to change established practice patterns, regardless of whether they be of clinicians, researchers, or funding agencies. It cannot reasonably be expected that many scientists will quickly discontinue practices in which they have been trained and become comfortable. It is also more efficient, and much more under one’s control, to continue to conduct efficacy studies without considering moderating variables or external validity because “the purpose is to study interventions under ideal conditions.” However, as illustrated above, this is only true if one does not intend to generalize one’s conclusions beyond the very limited sample and conditions of a given study,^{1,31} which is hardly ever the case in health promotion research.

There is an increasingly well-documented disparity between the large amount of information on efficacy and the very small amount of information on effectiveness and representativeness.^{21,22,40} To produce significant improvement in the current state of affairs, changes will be necessary on the part of researchers, funding organizations, journal reviewers, and grant review panels. We propose 4 specific changes—2 of which focus on researchers, 1 on journal editors, and 1 on funding organizations.

1. *Researchers should pay increased attention to moderating factors in both efficacy and effectiveness research.* Table 2 outlines how data collection and information about moderating factors, such as participant characteristics (reach) and setting characteristics (adoption), can be incorporated into both efficacy and effectiveness research in a manner appropriate to that phase. Using the RE-AIM framework, we suggest that researchers consider the types of settings, intervention agents, and individuals that they wish their program to be used by when designing and evaluating interventions. During efficacy studies, purposeful or oversam-

pling strategies can be used to include both specific end-user groups (e.g., minorities, less educated) and settings of interest. A critical concern for broader application—and an integral part of Flay’s original description¹⁶—was measurement of potential harmful outcomes. This part of his definition has seldom been addressed, but it needs to be.

Participatory research methods, including developing one’s intervention ideas collaboratively with members of the intended audience (individuals, intervention agents, and organization decisionmakers) should not be left for later phases of research but built into efficacy studies. More formal measures of adoption and setting level maintenance may need to wait until later effectiveness studies (Table 2), but both qualitative and quantitative “proxy measures” of these factors can and should be addressed in efficacy studies. Such information can lead to better tailoring of interventions to organizational culture in the same way that tailoring of intervention at the individual level has led to increased success.^{41,42} A final recommendation for both efficacy and effectiveness studies is to include a variety of intervention agents, to describe their backgrounds and levels of experience/expertise with regard to the target behavior, and to report on potential differences in implementation and outcomes associated with these differences.⁴³

As illustrated in Table 2, issues pertaining to moderating factors—and eventual translation into practice—are best addressed during the *planning* phases of research. RE-AIM, or other evaluation models,^{13,16} can be used to help plan and select samples, interventions, settings, and agents in ways that make it more likely that results will be replicated in later studies.

2. *Realize that public health impact involves more than just efficacy.* Our training and current review criteria all emphasize producing large effect sizes under tightly controlled conditions. To make a real-world impact, several other criteria are also necessary.

- a. At the individual level, several research groups have proposed that $\text{Impact} = \text{Reach (R)} \times \text{Efficacy (E)}$.^{44–47} It is not enough to produce a highly efficacious intervention. To have broad public health impact, an interven-