

EVALUATING THE RELEVANCE, GENERALIZATION, AND APPLICABILITY OF RESEARCH

Issues in External Validation
and Translation Methodology

LAWRENCE W. GREEN

University of California at San Francisco

RUSSELL E. GLASGOW

Kaiser Permanente Colorado

Starting with the proposition that "if we want more evidence-based practice, we need more practice-based evidence," this article (a) offers questions and guides that practitioners, program planners, and policy makers can use to determine the applicability of evidence to situations and populations other than those in which the evidence was produced (generalizability), (b) suggests criteria that reviewers can use to evaluate external validity and potential for generalization, and (c) recommends procedures that practitioners and program planners can use to adapt evidence-based interventions and integrate them with evidence on the population and setting characteristics, theory, and experience into locally appropriate programs. The development and application in tandem of such questions, guides, criteria, and procedures can be a step toward increasing the relevance of research for decision making and should support the creation and reporting of more practice-based research having high external validity.

Keywords: *evaluation; external validity; application; relevance; practice-based research; translation; dissemination; research methods*

AUTHORS' NOTE: This work was developed, in part, within a research collaboration on complex interventions funded by the Canadian Institutes of Health Research, and the Comprehensive Cancer Center at the University of California at San Francisco and, in part, from Grant #CA 90974-01 from the National Cancer Institute. We are indebted also to Barbara McCray of Kaiser Permanente of Colorado for her expert assistance with formatting, editing, and references.

EVALUATION & THE HEALTH PROFESSIONS, Vol. 29 No. 1, March 2006 126-153

DOI: 10.1177/0163278705284445

© 2006 Sage Publications

Recent developments in evidence-based medicine and public health guidelines have made the gap between science and practice more salient and embarrassing to the health professions and their sponsoring organizations (Institute of Medicine & Committee on Quality of Health Care in America, 2001; McGlynn et al., 2003). Meta-analyses and structured reviews that have produced the guidelines for practice from cumulative bodies of research literature have made it difficult to ignore the strength of evidence (e.g., direct evidence from randomized controlled trials [RCTs]) or specific practices on one hand, while unveiling on the other the limitations of that evidence (the weight of evidence) in its relevance for many practice situations. *Weight of evidence* refers to indirect evidence including non-experimental data, practitioner experiences, and the cumulative wisdom derived from systematic analysis of these and an understanding of the situations and populations in which they would be applied (e.g., Pasick, Hiatt, & Paskett, 2004). Much of the research on which practice guidelines have been based in the health professions has been strong on internal validity that provides strength of evidence extended from Type 1 translation traditions (Ames & McBride, in press), thanks to the emphasis that has been given to experimental control in the evaluation of evidence. These studies have been weak, however, on external validity that would add to the weight of evidence as applied to Type 2 translation of science to the varied circumstances of practice. Most judicial and regulatory agencies must rest their decisions more on weight of evidence because no single study involving human behavior or social change can unequivocally establish causation (Haack, 2005; Krimsky, 2005; Rohrbach, Grana, & Valente, in press; Steinberg & Luce, 2005). This commentary on the evidence-based practice literature examines the relative neglect of external validity and its consequences for the relevance, generalizability, and applicability of research in typical and varied circumstances of medical and public health practice. To be clear: Well-controlled efficacy studies have an important place in determining causation; the problem is that the current evidence base and evaluation schemes consist almost entirely of such research and very little “effectiveness” research (Flay, 1986) that attempts to study programs under typical, rather than optimal conditions (Glasgow, Lichtenstein, & Marcus, 2003).

Two major conclusions emerge from our observations. One is that some of the energy and resources of the evidence-based practice

movement needs to be directed toward development and application of criteria and measures of external validity. The second is that if the health professions and their sponsors want more widespread and consistent evidence-based practice, they will need to find ways to generate more practice-based evidence that explicitly addresses external validity and local realities (Green & Ottoson, 2004). Practice-based research would produce evidence that more accurately and representatively reflects the “program-context interactions” (Hawe, Shiell, Riley, & Gold, 2004) and circumstances in which the results of the research are expected to be applied. It would do so, of course, with some trade-off of the experimental control exercised in academically based research.

THE IMBALANCE IN INTERNAL AND EXTERNAL VALIDITY

A time-honored, well-developed, and widely accepted tradition of judging and rating internal validity has transcended the disciplines. The health professions have internalized the classical five criteria of Bradford Hill (Hill, 1965), which were based on Koch’s postulates (Koch, 1882) for proof of causation in biological studies from the 19th century. These have been reflected more widely across social service professions, building not just on the biomedical traditions but also agricultural and educational research where experimentation predated much of the action research in social and behavioral sciences. Campbell and Stanley’s (1963) widely used set of “threats to internal validity” were accompanied by seldom referenced “threats to external validity.” The focus on internal validity was justified on the grounds that without internal validity, external validity or generalizability would be irrelevant or misleading, if not impossible. The rating schemes of the Canadian Task Force on the Periodic Health Examination (1979), adopted also by the U.S. Preventive Services Task Force (1989) concerned themselves almost exclusively with internal validity. The greater weight given to evidence based on multiple studies than a single study was the main nod to external validity; however, even that was justified more on grounds of replicating the results in similar populations and settings than of representing different populations and settings. The criteria were adapted by the Community

Preventive Services Task Force, but with greater concern for external validity in recognition of the more varied public health circumstances of practice than clinical variations (Briss, Brownson, Fielding, & Zaza, 2004; Briss et al., 2000; Green & Kreuter, 2000). Similarly, reporting standards related to the CONSORT criteria (Mohrer, Schulz, Altman, & Lepage, 2001), used by the vast majority of medical and health promotion publications, focus predominantly on internal validity. Finally, numerous textbooks on research quality have tended to concern themselves primarily with designs for efficacy studies rather than effectiveness studies, although the growing field of evaluation has necessarily given more attention to issues of practice-based, real-time, ordinary settings (Glasgow, Klesges, Dzewaltowski, Estabrooks, & Vogt, in press; Green & Lewis, 1984). The use of the evaluation literature, however, could be strengthened by requiring systematic reviews and research syntheses to weigh the wider range of relevant evidence, not just the strongest controlled evaluations, in drawing inferences about generalizability. It could also be improved in its external validity with registries or repositories of evaluations conducted more routinely in more representative settings and populations. Finally, the application of evidence based on the strength of evidence and the weight of evidence, could be improved if there were guidelines for practitioners and decision makers for applying evidence.

With few exceptions (Cronbach, Glesser, Nanda, & Rajaratnam, 1972; Green, 2001; Leviton, 2001; Shadish, Cook, & Campbell, 2002), the evidence-based health practice literature seems to have lost focus on external validity. The irony of this seems lost on many of those who wonder why science has such difficulty achieving application and widespread adoption of evidence-based practice (EBP, a term generally attributed to Archie Cochrane and the Cochrane Collaboration [2004], derivative of their earlier emphasis on evidence-based medicine). The health field has been the most assiduous in its emphasis on EBP and internal validity (e.g., CONSORT, patient outcomes reporting trial [PORT], National Guideline Clearinghouse; *National Public Health Performance Standards* Program, n.d.; Substance Abuse and Mental Health Services Administration [SAMHSA], 2005). However, similar insistence on evidence-based programs has now been proposed with the usual emphasis on RCT designs by the Department of Education (www.eval.org/doe.fedreg.htm; response

by the American Evaluation Association: www.eval.org/doestatement.htm; response from the American Education Research Association: www.eval.org/doeaera.htm).

The purposes of this article are (a) to suggest a preliminary set of quality criteria for external validity; (b) to pose a set of questions to be asked of research evidence and guides to be used by practitioners, policy makers, and others involved in making decisions about applicability of research studies to their environment, practice, or population; and (c) to offer a series of steps in adapting evidence and incorporating it more systematically with theory, promising practices from related experiences in similar settings, and indigenous wisdom of those with firsthand experience in the setting.

WHAT ARE THE QUALITY ISSUES RELATED TO EXTERNAL VALIDITY IN EFFECTIVENESS AND DISSEMINATION RESEARCH?

GENERALIZATION THEORY

Cronbach et al. (1972), in their seminal book on generalizability theory, identified different facets across which program effects could be evaluated. They termed these facets *units* (e.g., individual patients, moderator variables, subpopulations), *treatments* (variations in treatment delivery or modality), *occasions* (e.g., patterns of maintenance or relapse over time in response to treatments), and *settings* (e.g., medical clinics, worksites, schools in which programs are evaluated), summarized as “utoS.” Table 1 lists the key components of Cronbach et al.’s generalization theory, and how it relates to more recent Type 2 translation frameworks. This theory also introduced concepts of robustness, or consistency of effects, across various domains, and of replication as an important criterion of strength of evidence. Although Cronbach et al. provided mathematical models for evaluating generalizability, neither these formulas nor the theory received much attention until recently when Shadish et al. (2002) incorporated many of these concepts into their conceptualization of external validity and causal generalizations. The Shadish et al. approach uses this theory as the basis to frame and discuss the strengths and limitations of various experimental (randomized) and quasi-experimental designs (this

TABLE 1
Relationship Among Different Approaches
to Evaluation of Generalizability

<i>Generalizability Theory</i> (Cronbach, Glesser, Nanda, & Rajaratnam, 1972)	<i>PCTs</i> (Tunis, Stryer, & Clancey, 2003)	<i>RE-AIM framework</i> (Glasgow, Klesges, Dziewaltowski, Bull, & Estabrooks, 2004, Table 2)
Units (u)	1. Representative participants	Reach (individual level) Participation rate Representativeness
Treatments (t)	2. Investigational interventions and standard of care	RE-AIM framework evaluates single and multicomponent programs and policies
Occasions (o)	3. Outcomes across time that are important to clinicians, decision makers, and consumers	Effectiveness (individual level) Effect size Adverse impacts Differential subpopulation response Maintenance (individual level) Individual: sustained treatment response
Settings (S)	4. Multiple settings	Adoption (setting & organization levels) Participation rate Representativeness
—	—	Implementation (setting & organization levels) Program component delivery Consistent component delivery
—	—	Maintenance (setting level) Setting: sustained program effectiveness and adaption over time

NOTE: PCTs = practical clinical trials; RE-AIM = reach, effectiveness, adoption, implementation, and maintenance.

reference is highly recommended for a thoughtful discussion of the pros and cons of randomization vs. alternative procedures for establishing experimental or statistical control of potential confounding variables). The Shadish et al. book is also an excellent example of balanced emphasis on external and internal validity.

ROBUSTNESS—OR BREADTH OF APPLICATION

Many of the above issues concern the similarity or dissimilarity of patients, conditions, intervention procedures, settings, and delivery

characteristics of those in a study to the broader population. This is precisely the issue that Tunis, Stryer, and Clancy (2003) were concerned about in their article on "practical clinical trials." They argued that many practitioners, organizational decision makers (e.g., purchasers), policy makers, and consumers do not find much of the evidence base from highly controlled randomized efficacy trials to be very relevant to their situation or the concerns that they have. Some may be lulled by the demand for EBP into a sense that local needs are less important than a cursory linking of apparent health problems to cookbook remedies (Hawe, 1996). To remedy these mismatches of evidence and the needs and circumstances that prevail in the "real world," Tunis et al. (2003) recommend conducting "practical trials" that have the characteristics in Table 1. In particular, they called for assessing outcomes important to decision makers (e.g., cost-effectiveness, quality of life), and using representative (or at least heterogeneous) samples of patients and settings. Finally, they recommended evaluating new treatments against realistic alternative interventions rather than no treatment or placebo controls.

The Tunis et al. article (2003) discussed important design and measurement elements but did not provide any methods or metrics to evaluate the extent to which a study meets their recommendations. The RE-AIM (reach, effectiveness, adoption, implementation, and maintenance) framework of Glasgow and colleagues (Glasgow, 2002; Glasgow, Vogt, & Boles, 1999; www.re-aim.org) is intended to aid the planning, conduct, evaluation, and reporting of studies having the goal of translating research into practice (Dzewaltowski, Estabrooks, & Glasgow, 2004; Klesges, Estabrooks, Glasgow, & Dzewaltowski, 2005). Table 1 illustrates how the RE-AIM framework relates to generalizability theory and to the practical clinical trials model. Table 2 provides definitions and evaluation questions related to the RE-AIM dimensions, each of which is generally assessed on a 0% to 100% scale.

Reach is a function of the participation rate and the representativeness of participants (Glasgow, Klesges, et al., in press). Effectiveness also has multiple components including the median effect size on primary outcome(s) (median rather than the mean is used to mitigate the impact of outliers given the small number of outcomes usually measured, to avoid undue influence of extreme values); any adverse impacts on quality of life or other outcomes; and differential impact

TABLE 2
RE-AIM Definitions and Questions To Ask to Assess Applicability (www.re-aim.org)

<i>RE-AIM Dimension</i>	<i>Definition</i>	<i>Questions to Ask</i>
Reach (individual level)	Participation rate among intended audience and representativeness of these participants	What percentage of the target population came into contact with or began program? Did program reach those most in need? Were participants representative of your practice setting?
Effectiveness (individual level)	Impact on key outcomes and quality of life Consistency of effects across subgroups	Did program achieve key targeted outcomes? Did it produce unintended adverse consequences? How did it affect quality of life? What did program cost as implemented and what would it cost in your setting?
Adoption (setting and/or organizational level)	Participation rate and representativeness of settings in the evaluation	Did low-resource organizations serving high-risk populations use it? Did program help the organization address its primary mission? Is program consistent with your values and priorities?
Implementation (setting and/or organizational level)	Level and consistency of delivery across program components and different staff members	How many staff members delivered the program? Did different levels of staff implement the program successfully? Were different program components delivered as intended?
Maintenance (individual and setting levels)	At individual level: Long-term effectiveness At setting level: Sustainability and adaptation of program	Did program produce lasting effects at individual level? Did organizations sustain the program over time? How did the program evolve? Did those persons and settings that showed maintenance include those most in need?

across population subgroups (Glasgow, Klesges, et al., in press), with special reference to groups identified in health disparities research (Institute of Medicine, 2003). The RE-AIM framework considers results not only at the individual level but also at the setting level. Adoption is a function of the participation rate among settings (e.g., organizations, clinics, schools) and the representativeness of these settings (e.g., do low resource and rural settings participate in rates equal to other settings?). Implementation includes the median level of delivery of different components of an intervention, and consistency of delivery across implementation staff (Glasgow, Nelson, Strycker, & King, in press). Finally, maintenance refers to long-term effectiveness at the individual level and to sustainability of a program at the setting or organizational level.

The RE-AIM model, adapted and expanded from earlier work on diffusion theory (Rogers, 2003) and health promotion planning (Green & Kreuter, 1991), has been used with increasing frequency in recent years to frame evaluation questions and to report on translation issues (see www.re-aim.org/publications; Eakin, Bull, Glasgow, & Mason, 2002; Will, Farris, Sanders, Stockmyer, & Finkelstein, 2004). It has also been helpful in identifying existing gaps in the health promotion evidence base (Glasgow, Klesges, Dzewaltowski, Bull, & Estabrooks, 2004; Estabrooks, Dzewaltowski, Glasgow, & Klesges, 2002). At a conceptual level, it is becoming widely accepted that at the individual level intervention impact is a function of reach multiplied by effectiveness (Abrams et al., 1996; Prochaska, Velicer, Fava, Rossi, & Tsoh, 2001; Glasgow, Klesges, et al., in press). However, it is not entirely straightforward how to form comprehensive indices of either reach or effectiveness, each of which is composed of multiple elements. While beyond the scope of this article, Glasgow, Klesges, et al. (in press) proposed specific procedures to form summary indices of individual level as well as "setting level" impact, effectiveness, and efficiency (Glasgow, Nelson, et al., in press; Green & Kreuter, 2005). Table 1 summarizes how the various models discussed thus far approach the various issues involved in external validity.

PROGRAM ADAPTATION AND EVOLUTION

A final challenging issue related to dissemination and external validity concerns the adaptability of programs. Program developers,

especially if they become vendors of the programs or of training or materials to accompany the program, are often primarily concerned with the “fidelity” (Bellg et al., 2004) with which their intervention protocols are translated into practice. There is merit to this concern, as there is likely some level beyond which modifications and adaptations to a protocol result in a program that no longer closely resembles the original evidence-based protocol and may not be effective. On the other hand, it is well established that program adopters seldom adopt or implement a program exactly as it was originally tested. Rather, there is some degree of reinvention or customization that occurs (Rogers, 2003). From the perspective of community-based participatory approaches (Israel, Eng, Schulz, & Parker, 2005), this is not only pervasive but also generally desirable. Where is the balance between these two opposing criteria of complete fidelity and complete adaptation or customization to local settings, clientele, resources, and priorities? There is presently no consensus on this issue (see, e.g., the debates surrounding the national evaluation of the Fighting Back community programs in substance abuse prevention; Green & Kreuter, 2002); however, we suggest that the solution may lie in the specification and documentation of (a) a limited set of key components or principles of an evidence-based program (Ory, Evashwick, Glasgow, & Sharkey, 2005), (b) the range of permissible adaptations that still retains the essential elements of the original efficacy-tested intervention (Castro, Barrera, & Martinez, 2004), and (c) justifications of theory-driven and experience-driven deviations (e.g., weight of evidence) from evidence-based recommendations, as related to moderating variables and history in the local situation. A given adaptation could then be rated to the extent that it implemented such key components and made “appropriate” adaptations versus those of unknown or nonrecommended methods. These principles apply most comfortably when the intervention is a discreet, contained service or professional action, such as a medicine, a vaccine, or a specific message as part of a counseling session. They become more difficult to apply when the intervention is a complex program made up of many discrete interventions, such as the full range of interventions required to predispose, enable, and reinforce a set of behavioral and environmental determinants of a specific health outcome (as in the PRECEDE-PROCEED model; Green & Kreuter, 2005; or the chronic illness care model; Glasgow, Orleans, Wagner, Curry, & Solberg,

2001; Wagner, 1998). These difficulties are addressed in the final sections of this article with specific elaborations on the adaptation process using theory, expert opinion, and local participation in the process. We seek, in short, a “best process” of program planning to complement “best practices.”

EXTERNAL VALIDITY QUALITY RATING CRITERIA

To help address the relative dearth of criteria and standards related to external validity and potential for implementation, we propose a set of six specific ratings, under the three headings of reach and representativeness, implementation and consistency of effects, and maintenance and institutionalization (see Table 3). We further recommend that these criteria, or a similar set of quality ratings, be added to or used in addition to existing guidelines and rating scales such as CONSORT (Mohrer et al., 2001); critique by Gross, Mallory, Heiat, & Krumholz (2002); TREND (Des Jarlais, Lyles, Crepaz, & TREND Group, 2004); critique by Dzewaltowski, Estabrooks, Klesges, & Glasgow (2004); the Jadad scale (Jadad et al., 1996); and used by review groups such as AHRQ Evidence-Based Practice Centers (Agency for Health Research and Quality, 2005); Cochrane reviewers (Jackson, Waters, & The Guidelines for Systematic Reviews, 2004); the U.S. Preventive Services Task Force (1989, 1996) and the Community Preventive Services Guides reviewers (Briss et al., 2004). To our knowledge, only the Community Guide (Truman et al., 2000; Zaza et al. & Task Force on Community Preventive Services, 2000) currently considers many of these issues, and their external validity criteria are often necessarily subjective in the absence of specific criteria such as those suggested below.

REACH AND REPRESENTATIVENESS CRITERIA

1. Participation: Are there analyses of the participation rate among potential (a) settings, (b) delivery staff, and (c) patients (consumers)? These criteria provide a rough index of the potential public health or larger population impact of a program when it is taken to scale, assuming that eligibility and exclusion criteria are specified at each of these levels.

TABLE 3
Proposed Quality Rating Criteria for External Validity

I.	Reach and representativeness
A.	Participation: Are there analyses of the participation rate among potential (a) settings, (b) delivery staff, and (c) patients (consumers)?
B.	Target audience: Is the intended target audience stated for adoption (at the intended settings such as worksites, medical offices, etc.) and application (at the individual level)?
C.	Representativeness—Settings: Are comparisons made of the similarity of settings in study to the intended target audience of program settings—or to those settings that decline to participate?
D.	Representativeness—Individuals: Are analyses conducted of the similarity and differences between patients, consumers, or other subjects who participate versus either those who decline, or the intended target audience?
II.	Program or policy implementation and adaptation
A.	Consistent implementation: Are data presented on level and quality of implementation of different program components?
B.	Staff expertise: Are data presented on the level of training or experience required to deliver the program or quality of implementation by different types of staff?
C.	Program adaptation: Is information reported on the extent to which different settings modified or adapted the program to fit their setting?
D.	Mechanisms: Are data reported on the process(es) or mediating variables through which the program or policy achieved its effects?
III.	Outcomes for decision making
A.	Significance: Are outcomes reported in a way that can be compared to either clinical guidelines or public health goals?
B.	Adverse consequences: Do the outcomes reported include quality of life or potential negative outcomes?
C.	Moderators: Are there any analyses of moderator effects—including of different subgroups of participants and types of intervention staff—to assess robustness versus specificity of effects?
D.	Sensitivity: Are there any sensitivity analyses to assess dose-response effects, threshold level, or point of diminishing returns on the resources expended?
E.	Costs: Are data on the costs presented? If so, are standard economic or accounting methods used to fully account for costs?
IV.	Maintenance and institutionalization
A.	Long-term effects: Are data reported on longer term effects, at least 12 months following treatment?
B.	Institutionalization: Are data reported on the sustainability (or reinvention or evolution) of program implementation at least 12 months after the formal evaluation?
C.	Attrition: Are data on attrition by condition reported, and are analyses conducted of the representativeness of those who drop out?

2. Target audience: Is the intended target audience stated for adoption (at the intended settings such as worksites, medical offices, etc.) and application (at the individual level)? Without an explicit statement of the intended target audience(s), it is difficult, if not impossible, to evaluate the representativeness of participants in a study. If the targets are stated, then the next two questions should be answered.
3. Representativeness—Settings: Are comparisons made of the similarity of settings in the study to the intended target audience of program settings—or to those settings whose authorities declined to have their settings included in the current study?
4. Representativeness—Individuals: Are analyses conducted of the similarity and differences between patients, consumers, or other respondents who participate versus either those who decline or the intended target audience?

These measures of representativeness allow evaluation of the extent to which those most in need of program services are included in the studies of the program. Although there are a large number of potential characteristics on which study participants can be compared to other samples, we recommend the following as a feasible and relatively low burden set of variables: age, gender, education, income, race and ethnicity, number or type of medical conditions, health literacy, and status on the particular condition or problem being studied (it is recognized that the last two variables may be more challenging to collect but are recommended because of their established relevance to outcomes (Institute of Medicine, 1999, 2004). At the setting level, characteristics to be reported vary depending on the type of program (e.g., worksite vs. health clinic vs. school) but should include size, urban versus rural setting, availability of needed resources, and level of need of clients served (or type of employees), and strength of the commitment of management to the program. Depending on the intended range of generalization or exportation, ratings might also be added on program service array, linkage to other health or human services that a program does not offer, and fiscal environment (e.g., payer mix, local economy). Even practice-based research networks supported by innovative federal and foundation funding to address some of the issues of generalizability have come under criticism for their representativeness of the universe of local, state, and regional practice settings (Norquist, 2001).

IMPLEMENTATION AND CONSISTENCY OF EFFECTS

1. Program or policy implementation and adaptation
 - a. Consistent implementation: Are data presented on the level and quality of implementation of major intervention components?
 - b. Staff expertise: Are data presented on the level of training or experience required to deliver the program, or on the quality of implementation by different types of staff?
 - c. Program adaptation: Is information reported on the extent to which different settings modified or adapted the program to fit different types of population groups or individuals (Brook & Lohr, 1985)?
 - d. Mechanisms: Are data reported on the process(es) or mediating variables through which the program or policy achieved its effects?

One of the most common reasons that programs fail when applied in community settings is that they are not implemented with the same level of skill or consistency as in the controlled trials documenting program efficacy (Basch, Sliepecevic, & Gold, 1985). There are many reasons for this including the fact that in efficacy studies the intervention staff often have unusually high levels of training, expertise, or supervision, or they are employed solely to deliver the intervention being evaluated rather than having multiple competing responsibilities (Stange, Woolf, & Gjeltrema, 2002). Therefore, it is important to document the extent to which different program components are delivered, and the level of training or skill required to implement the program successfully.

We purposefully use the term *implementation* rather than *fidelity* (Bellg et al., 2004) to communicate that modifications to a protocol may be either problematic or advantageous. The issue of program adaptation or customization to fit local needs, situations, and preferences is discussed below; however, when reporting on a previously developed program, it is important to document how the program was modified and evolved over time (Rotheram-Borus & Flannery, 2004). In community settings, a program is almost always adapted or “re-invented” to address local concerns and resources.

2. Outcomes for decision making:
 - a. Significance: Are outcomes reported in a way that can be compared to either clinical guidelines (Tinetti, Bogardus, & Agostini, 2004; Walter, Davidowitz, Heineken, & Covinsky, 2004) or public

health goals or guidelines (U.S. Department of Health and Human Services, 2000; Briss, Brownson, Fielding, & Zaza, 2004)?

- b. Adverse consequences: Do the outcomes reported include quality-of-life or potential negative outcomes?
- c. Moderators: Are there any analyses of moderator effects—including of different types of intervention staff—to assess robustness versus specificity of effects?
- d. Sensitivity: Are there any sensitivity analyses to assess dose-response effects, threshold level, or point of diminishing returns on the resources expended?

The recommended criteria above allow readers to go beyond the primary outcomes of a particular study to evaluate the potential public health relevance of program results, especially when combined with information on program reach and adoption. Quality of life (Kaplan, 2003) is a measure of ultimate impact and can provide a common metric for comparing different programs for different target problems by comparing their impact on health-related quality of life (Ware & Kosinski, 2001). As documented in the Institute of Medicine report (1999) on medical errors, there are also often unintended consequences of health care interventions. One of the most likely unintended consequences of health promotion interventions focused on a given issue (e.g., stopping smoking) may be negative impacts on or decreased attention to other important health behaviors (e.g., obesity rates or eating patterns). Finally, the issue of moderator effects is important to determine if programs are successful with segments of the population most in need of assistance, and whether the characteristics of these priority target populations moderate the relationship between the tested intervention and the outcomes. In particular, to achieve national goals of reducing or eliminating health disparities, we need to evaluate program impact along these and related dimensions such as health literacy.

- e. Realistic cost: Are data on the costs presented? If so, are standard economic or accounting methods used to fully account for costs?

Economic issues, including cost, are some of the first questions that potential program adoptees have when considering new alternatives, and are of vital importance to decision-making bodies such as businesses, health departments, or Centers for Medicare and Medicaid. At

minimum, program costs should be reported, using standard accepted methods (Gold, Siegel, Russell, & Weinstein, 2003) so that programs are compared on a level playing field. We encourage reporting of more detailed economic outcomes such as cost-effectiveness or cost utility results but realize that such more sophisticated analyses may be beyond the resources of some projects.

MAINTENANCE AND INSTITUTIONALIZATION

1. Long-term effects: Are data reported on longer term effects, at least 12 months following treatment?
2. Institutionalization: Are data reported on the sustainability (or reinvention or evolution) of program implementation at least 12 months after the formal evaluation?
3. Attrition: Are data on attrition by condition reported, and are analyses conducted of the representativeness of those who dropout?

The literature indicates that programs that have initially large effects are not necessarily equally successful long term (Orleans, 2000) or institutionalized by the organizations conducting them. In addition, in community settings, many factors besides effect size affect decisions to continue a program following an initial trial. To have lasting public health benefit, programs need to have longer term benefits for participants, and to be continued over time by program sponsors.

These criteria are of value individually; however, the next step in elevating the importance of such external validity criteria should be to evaluate the use of all 16 items as a scale, and to demonstrate the reliability and usefulness of the scale. External validity of studies purported to guide best practices is partly a matter of generalizability to many situations and populations (also referred to above as *robustness*). External validity for the practitioner or program planner who would adopt the practice recommended by previous research, however, is conversely a matter of its particular relevance to the local setting, population, and circumstances. Generalizability or robustness often does not encompass a particular combination of population, setting, and circumstances. The impossibility of ever having sufficient numbers of studies to cover all the combinations of settings, populations, and circumstances calls on theory, experience, and local data and wisdom to fill the gaps in external evidence. The blending of these into a sensible interpretation of evidence for one's local purposes is a

combination of science and the art of practice but can be made more systematic by the strategic combination of experimental evidence with local surveillance evidence, theory, professional judgment, and participatory planning with those who have local experience.

ALIGNMENT OF PRIORITY DETERMINANTS WITH PROGRAM COMPONENTS

Two levels of alignment are suggested by the RE-AIM model (Glasgow et al., 2004) and by the PRECEDE-PROCEED model (Green & Kreuter, 2005). *PRECEDE* refers to predisposing, reinforcing, and enabling constructs in ecological diagnosis and evaluation, and *PROCEED* refers to policy, regulatory, and organizational constructs in educational and ecological development. It is a generic logic model suggesting causal priorities for the focus of diagnostic baseline studies for planning and evaluating programs, and a procedural model for the specific order of assessments that should precede the selection of interventions and the alignment of interventions with the ecological levels of organization. Figure 1 suggests the levels of intervention, and Figure 2 illustrates the relationships among the various types of evidence and theory used to complement and fill gaps in the evidence derived from more or less generalizable experimental evidence from other places.

One level of alignment is at the institutional or organizational level of adoption of the intervention, the other is at the individual level of implementation. Program components must be aligned with levels of policy, regulatory, or organizational change needed from groups of individuals representing organizations or whole communities (which may be states, provinces, or even countries). This is an *ecological alignment*. It sets the stage for putting the program into the broader environmental context in which the change must occur, and it recognizes the interdependence of levels in a social system. Each subsystem (such as a group of practitioners or a family) relates to a larger system (such as an organization or community), and each system depends for its maintenance on multiple subsystems. The program components at the higher levels might be changing the nonsmoking policies in a building, the food choices in vending machines of a school, and the carpooling lanes on a commuter's highway. These examples of

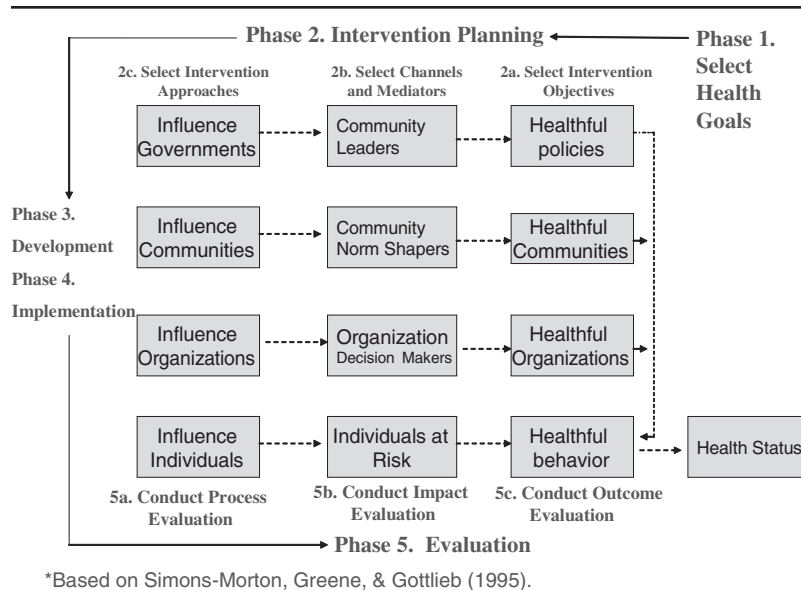


Figure 1: Levels of Intervention Constituting Programs to Effect More Complex Behavioral, Environmental, or Social Changes in Support of Specific Health Outcomes in Populations (Adapted From Simons-Morton et al., 1989; Simons-Morton, Parcel, & O'Hara, 1988; as adapted for Green & Kreuter, 2005)

organizational and environmental changes can have secondary or ecological effects on the behavior of large numbers of people without necessarily trying to persuade their change through direct communication. These elements of a program operate primarily through enabling factors for environmental change to predispose, enable, and reinforce, in turn, behavioral change in whole populations, or reduced exposure to environmental risks that could have a direct effect on the health of the population. This contextualization process is one of the tasks of putting evidence from another setting or environment into the local setting and population circumstances.

Evidence on such ecological interventions is less likely to pass the internal validity screen in systematic reviews that lead to "best practice" guidelines because they can seldom be studied with randomized designs. The evidence of the effectiveness of such broad, ecological interventions, however, will tend to have greater external validity and be more persuasive to planners and policy makers, insofar as they achieve similar effects in multiple jurisdictions and have more pervasive

*Uses of Evidence & Theory in Population-Based,
Diagnostic, Planning, & Evaluation Models*

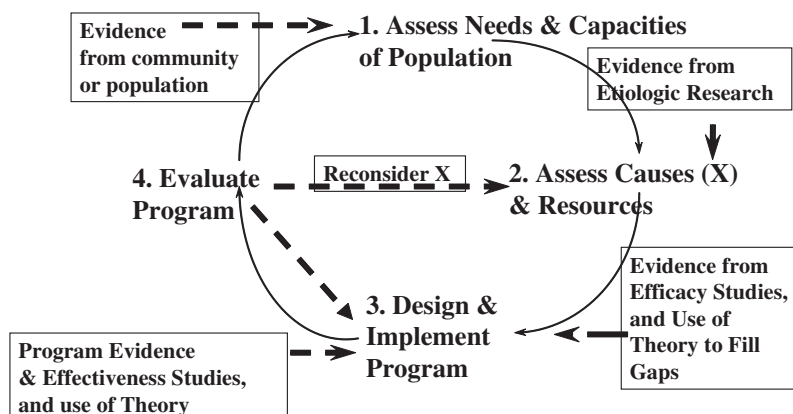


Figure 2: Sequence of Planning Process in Which Evidence From Various Sources Are Combined to Achieve Optimum Relevance to the Local Setting and Population, Grounding in Evidence From More or Less Generalizable Research, and Continuous Evaluation of the Fit With Local Needs and Circumstances (Green & Kreuter, 2005, fig 5-1)

benefits (Hawe, Noort, King, & Jordens, 1997). The influence of the California and Massachusetts comprehensive tobacco control programs and some of their components such as the increased cigarette taxes are examples of widely acknowledged and emulated interventions that were more influential on the adoption of policies by other states than were prior attempts to mount more rigorously controlled community trials (Pentz, Jasuja, Rohrbach, Sussman, & Bardo, in press).

At the more individual, behavioral, family, or other microlevels of social systems, the task of alignment is between specific predisposing, enabling, or reinforcing factors and the more specific program components, interventions, or methods for which evidence of their effectiveness has been derived from previous research that has greater internal validity but less external validity. The need to bring theory, experience, and professional and community judgment to bear on interpreting evidence from afar as it pertains to local settings, populations, and circumstances arises particularly at this level.

INTERVENTION MATCHING, MAPPING, POOLING, AND PATCHING

Several alignments of evidence with circumstances and populations are required. The first (a) is to match the ecological levels of the target setting with evidence for needs and changeability for each level from community or organizational to practitioner or individual at risk, and evidence of variable rates of uptake for each major category of organization or individual in response to different interventions; (b) then using theories to map specific interventions from prior research and practice to specific predisposing, enabling, and reinforcing factors that are theorized to determine the needs and change process, and (c) a pooling of prior interventions and community-preferred interventions that might have less evidence to support them, but that might be needed to patch or fill gaps in the evidence-based “best practices.” The terms *match*, *map*, *pool*, and *patch* align roughly with the planning models or procedures that use these terms as acronyms or analogies (see Simons-Morton et al., 1988, for the MATCH model; Bartholomew, Parcel, Kok, & Gottlieb, 2001, for the Intervention Mapping process; D’Onofrio, 2001, and Sussman, 2001, for the pooling and warehousing process; and Green & Kreuter, 2005, pp. 203–204, for the PATCH model).

THE ECOLOGICAL LEVEL OF MATCHING

The ecological approach calls first for a matching of types of interventions with the level at which, or channels and settings through which, they can have their effects—community (including homes, restaurants, other public or commercial places, and mass media), schools, worksites, and health care institutions, and subsystem levels of informal groups—neighborhoods, families, and individuals. We use the term *matching* for this level of alignment because a cogent model for aligning interventions or program components with ecological levels is called MATCH, for Multilevel Approach to Community Health. It emerged as ecological approaches saw a renaissance in public health and health promotion, bringing renewed interest in environmental risk conditions after a period of highly focused research and development on individual risk factors (Green, Richard, & Potvin, 1996; McLeroy, Bibeau, Steckler, & Glanz, 1988; Stokols, 1992).

Mediating and Moderating Variables

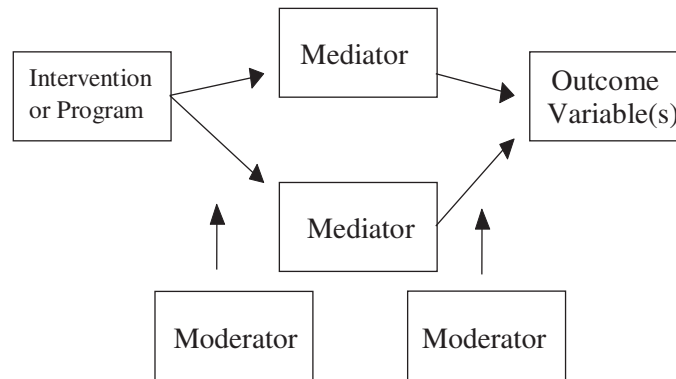


Figure 3: Mediating Variables, as the Causal, Intermediate Changes Through Which Interventions or Programs Can Affect Outcomes, Are Moderated in Their Response to Interventions and in Their Impact on Outcomes by the Characteristics of the Persons, Settings, or Circumstances of the Intervention

Whether the ecological moniker was real or metaphorical (Trickett & Mitchell, 1992), it has taken firm hold in the community intervention and public health fields.

THE USE OF THEORY TO MAP MEDIATING AND MODERATING VARIABLES

The empirical evidence will never cover all the combinations of interventions, mediating variables at which they are targeted to influence the desired outcomes, and moderating variables in the characteristics of the setting, population, and circumstances that influence the relationships between interventions, mediating and outcome variables (see Figure 3). Theory comes to the rescue of the program planner in mapping the evidence specific to the mediating links and the population and setting characteristics at hand and in filling gaps in the setting-specific evidence with extrapolations from similar settings and their population-problem-circumstance configurations (Poland & Green, 2000; Sussman, 2001). *Intervention mapping* was the term coined by Bartholomew et al. (2001) to identify a set of specific steps

the program planner can follow to get from evidence to intervention using theory to fill gaps in evidence, and to query the evidence for its relevance to the situation and population for which a program is being planned.

POOLING AND PATCHING PRIOR AND EXISTING INTERVENTIONS

Even in the absence of published evidence, a program-planning effort inevitably addresses a health problem that someone, somewhere, has tried to solve before. Other attempts elsewhere and in the very setting of the new planning effort should be considered a source of tacit knowledge to be reviewed and pooled from the best experience of prior attempts to address the problem. Professionals typically tap into their professional networks by telephone, meetings, and the Internet. D'Onofrio (2001) observed, however, that "to date, no systematic procedures have been suggested for accomplishing this task as part of the program-planning process" (p. 158). She presented a set of specific procedures for the planner or practitioner to follow in identifying prior interventions from which ideas, inspiration, and insight can be drawn, and bad ideas discarded, which she refers to as *pooling*. Schorr (1997) presented a case for more reliance on replication of model programs and less dependence on the plodding pace of randomized trials to educe "best practices" (pp. 60-64).

Existing programs and experience with related activities in a community can also be a source of even richer information than prior interventions conducted elsewhere because they are indigenous to the community or setting and were designed with the same population and more similar circumstances than most prior interventions. Here is where the PATCH (Planned Approach to Community Health) adaptation of PRECEDE-PROCEED offered a useful existing community programs and policies matrix and downloadable checklist available online (Centers for Disease Control, 2001).

CONCLUSIONS AND RECOMMENDATIONS

This article sought to identify some points of convergence of the "bottom-up" methods of planners and practitioners in judging the relevance of studies for their local situation and the "top-down" criteria

that apply to judging the external validity of studies at large, without reference to specific local situations. Some criteria on the two sides converge, such as noting the representativeness or the characteristics of respondents included in the studies used to infer "best practices." Other criteria do not converge, largely because the evidence available is necessarily limited and inevitably leaves local practitioners and program planners to fill the gaps in evidence. Some of these gaps are in relation to the ecological levels to which the evidence does and does not apply, and the setting and population characteristics and circumstances. For these gaps, the practitioner or program planner must turn (a) to theory to generalize from existing evidence in health and other fields to the local circumstances they face; (b) to experience of other practitioners and planners dealing with similar populations, problems, and circumstances; and (c) to indigenous wisdom of those who are stakeholders and have the best intuitive understanding and familiarity with the local population and circumstances.

We recommend, then, the continued development and formalization of the practitioner-planner procedures for reviewing and filling gaps in the evidence, and of the criteria for judging the generalizability or external validity of studies. These would not displace the criteria of internal validity that should continue to guide the metareviews of evidence as a first screen because without internal validity, there can be no external validity. However, with greater attention to the issues of external validity and practice-based research to enhance the relevance to particular settings, populations, and circumstances, the credibility of evidence-based "best practices" will grow, and the application and appropriate adaptation of them will lead to better programs and practice.

REFERENCES

- Abrams, D. B., Orleans, C. T., Niaura, R. S., Goldstein, M. G., Prochaska, J. O., & Velicer, W. (1996). Integrating individual and public health perspectives for treatment of tobacco dependence under managed health care: A combined stepped care and matching model. *Annals of Behavioral Medicine, 18*, 290-304.
- Agency for Health Research and Quality. (2005). *Evidence-based practice centers: Synthesizing scientific evidence to improve quality and effectiveness in health care*. Retrieved September 28, 2005, from www.ahrq.gov/clinic/epc

- Ames, S. L., & McBride, C. (in press). Involvement of genetics, neuroscience, and other basic research. *Evaluation & the Health Professions*.
- Bartholomew, L. K., Parcel, G. S., Kok, G., & Gottlieb, N. H. (2001). *Intervention mapping: Designing theory and evidence-based health promotion programs*. Mountain View, CA: Mayfield.
- Basch, C. E., Sliepcevich, E. M., & Gold, R. S. (1985). Avoiding Type III errors in health education program evaluations. *Health Education Quarterly*, 12, 315-331.
- Bellg, A. J., Borrelli, B., Resnick, B., Ogedegbe, G., Hecht, J., Ernst, D., et al. (2004). Enhancing treatment fidelity in health behavior change studies: Best practices and recommendations from the Behavior Change Consortium. *Health Psychology*, 23, 443-451.
- Briss, P. A., Brownson, R. C., Fielding, J. A., & Zaza, S. (2004). Developing and using the Guide to Preventive Health Services: Lessons learned about evidence-based public health. *Annual Review of Public Health*, 25, 281-302.
- Briss, P. A., Zaza, S., Pappaioanou, M., Fielding, J., Wright-de Agüero, L., Truman, B. I., et al. (2000). Developing an evidence-based guide to community preventive services—Methods. *Preventive Medicine*, 18, 35-43.
- Brook, R. H., & Lohr, K. N. (1985). Efficacy, effectiveness, variations, and quality. *Medical Care*, 23, 710-722.
- Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research on teaching*. Chicago: Rand McNally.
- Canadian Task Force on the Periodic Health Examination. (1979). The periodic health examination. *Canadian Medical Association Journal*, 121, 1193-1254.
- Castro, F. G., Barrera, M., Jr., & Martinez, C. R., Jr. (2004). The cultural adaptation of prevention interventions: Resolving tensions between fidelity and fit. *Prevention Science*, 5, 41-45.
- Centers for Disease Control. (2001). *Chronic disease prevention—PATCH*. Available from www.cdc.gov/nccdphp/patch/00binaries/PATCHCh5.pdf
- Cochrane Collaboration. (2004). Glossary. In Cochrane Library (Ed.), *Cochrane collaboration* (2nd ed.). Oxford, UK: Oxford University Press.
- Cronbach, L. H., Glesser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: John Wiley.
- Des Jarlais, D. C., Lyles, C., Crepaz, N., & TREND Group. (2004). Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: The TREND statement. *American Journal of Public Health*, 94, 361-366.
- D'Onofrio, C. N. (2001). Pooling information about prior interventions: A new program planning tool. In S. Sussman (Ed.), *Handbook of program development for health behavior* (pp. 158-203). Thousand Oaks, CA: Sage.
- Dzewaltowski, D. A., Estabrooks, P. A., & Glasgow, R. E. (2004). The future of physical activity behavior change research: What is needed to improve translation of research into health promotion practice? *Exercise and Sport Sciences Reviews*, 32, 57-63.
- Dzewaltowski, D. A., Estabrooks, P. A., Klesges, L. M., & Glasgow, R. E. (2004). TREND: An important step, but not enough [Letter to Editor]. *American Journal of Public Health*, 94, 1474.
- Eakin, E. G., Bull, S. S., Glasgow, R. E., & Mason, M. (2002). Reaching those most in need: A review of diabetes self-management interventions in disadvantaged populations. *Diabetes Metabolism Research and Reviews*, 18, 26-35.
- Estabrooks, P. A., Dzewaltowski, D. A., Glasgow, R. E., & Klesges, L. M. (2002). School-based health promotion: Issues related to translating research into practice. *Journal of School Health*, 73, 21-28.
- Flay, B. R. (1986). Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. *Preventive Medicine*, 15, 451-474.

- Glasgow, R. E. (2002). Evaluation of theory-based interventions: The RE-AIM model. In K. Glanz, F. M. Lewis, & B. K. Rimer (Eds.), *Health behavior and health education* (3rd ed., pp. 531-544). San Francisco: Jossey-Bass.
- Glasgow, R. E., Klesges, L. M., Dzewaltowski, D. A., Bull, S. S., & Estabrooks, P. (2004). The future of health behavior change research: What is needed to improve translation of research into health promotion practice? *Annals of Behavioral Medicine*, 27, 3-12.
- Glasgow, R. E., Klesges, L. M., Dzewaltowski, D. A., Estabrooks, P. A., & Vogt, T. M. (in press). Evaluating the overall impact of health promotion programs: Using the RE-AIM framework to form summary measures for decision making involving complex issues. *Health Education Research*.
- Glasgow, R. E., Lichtenstein, E., & Marcus, A. C. (2003). Why don't we see more translation of health promotion research to practice? Rethinking the efficacy to effectiveness transition. *American Journal of Public Health*, 93, 1261-1267.
- Glasgow, R. E., Nelson, C. C., Strycker, L. A., & King, D. K. (in press). Using RE-AIM metrics to evaluate diabetes self-management support interventions. *American Journal of Preventive Medicine*.
- Glasgow, R. E., Orleans, C. T., Wagner, E. H., Curry, S. J., & Solberg, L. I. (2001). Does the chronic care model serve also as a template for improving prevention? *Milbank Quarterly*, 79, 579-612.
- Glasgow, R. E., Vogt, T. M., & Boles, S. M. (1999). Evaluating the public health impact of health promotion interventions: The RE-AIM framework. *American Journal of Public Health*, 89, 1322-1327.
- Gold, M. R., Siegel, J. E., Russell, L. B., & Weinstein, M. C. (2003). *Cost-effectiveness in health and medicine*. New York: Oxford University Press.
- Green, L. W. (2001). From research to "best practices" in other settings and populations. *American Journal of Health Behavior*, 25, 165-178.
- Green, L. W., & Kreuter, M. W. (1991). *Health promotion planning: An education and environmental approach*. Mountain View, CA: Mayfield.
- Green, L. W., & Kreuter, M. W. (2000). Commentary on the emerging Guide to Community Preventive Services from a health promotion perspective. *American Journal of Preventive Medicine*, 18, 7-9.
- Green, L. W., & Kreuter, M. W. (2002). Commentary. Fighting back or fighting themselves: Community coalitions against substance abuse and their use of best practices. *American Journal of Preventive Medicine*, 23, 303-306.
- Green, L. W., & Kreuter, M. W. (2005). *Health promotion planning: An educational and ecological approach* (4th ed.). New York: McGraw-Hill.
- Green, L. W., & Lewis, F. M. (1984). *Measurement and evaluation in health education and health promotion*. Palo Alto, CA: Mayfield.
- Green, L. W., & Ottoson, J. M. (2004). From efficacy to effectiveness to community and back: Evidence-based practice vs. practice-based evidence. In *Proceedings from conference: From Clinical Trials to Community: The Science of Translating Diabetes and Obesity Research*. Bethesda, MD: National Institutes of Diabetes, Digestive and Kidney Diseases, National Institutes of Health.
- Green, L. W., Richard, L., & Potvin, L. (1996). Ecological foundations of health promotion. *American Journal of Health Promotion*, 10, 270-281.
- Gross, C. P., Mallory, R., Heiat, A., & Krumholz, H. M. (2002). Reporting the recruitment process in clinical trials: Who are these patients and how did they get there? *Annals of Internal Medicine*, 137, 10-16.
- Haack, S. (2005). Trial and error: The Supreme Court's philosophy of science. *American Journal of Public Health*, 95, S66-S73.

- Hawe, P. (1996). Needs assessment must become more needs focused. *Australian and New Zealand Journal of Public Health*, 20, 473-478.
- Hawe, P., Noort, M., King, L., & Jordens, C. (1997). Multiplying health gains: The critical role of capacity-building within health promotion programs. *Health Policy*, 39, 29-42.
- Hawe, P., Shiell, A., Riley, T., & Gold, L. (2004). Methods for exploring implementation variation and local context within a cluster randomised community intervention trial. *Journal of Epidemiology and Community Health*, 58, 788-793.
- Hill, A. B. (1965). The environment and disease: Association or causation. *Proceedings of the Royal Society of Medicine*, 58, 295-300.
- Institute of Medicine. (1999). *To err is human: Building a safer health system*. Washington, DC: National Academies Press.
- Institute of Medicine. (2003). *Unequal treatment: Confronting racial and ethnic disparities in health care*. Washington, DC: National Academies Press.
- Institute of Medicine. (2004). *Health literacy: A prescription to end confusion*. Washington, DC: National Academies Press.
- Institute of Medicine & Committee on Quality of Health Care in America (2001). *Crossing the quality chasm: A new health system for the 21st Century*. Washington, DC: National Academies Press.
- Israel, B. A., Eng, E., Schulz, A. J., & Parker, E. A. (2005). *Methods in community-based participatory research for health*. San Francisco: Jossey-Bass.
- Jackson, N., Waters, E., & The Guidelines for Systematic Reviews of Health Promotion and Public Health Interventions Task Force. (2004). The challenges of systematically reviewing public health interventions. *Journal of Public Health*, 26, 303-307.
- Jadad, A. R., Moore, R. A., Carroll, D., Jenkinson, C., Reynolds, D. J., Gavaghan, D. J., et al. (1996). Assessing the quality of reports of randomized clinical trials: Is blinding necessary? *Control Clinical Trials*, 17, 1-12.
- Kaplan, R. M. (2003). The significance of quality of life in health care. *Quality of Life Research*, 12, 3-16.
- Klesges, L. M., Estabrooks, P. A., Glasgow, R. E., & Dziewaltowski, D. (2005). Beginning with the application in mind: Designing and planning health behavior change interventions to enhance dissemination. *Annals of Behavioral Medicine*, 29, 66S-75S.
- Koch, R. (1882). *Classifying infectious diseases: The etiology of tuberculosis*. Berlin, Germany: Robert Koch.
- Krimsky, S. (2005). The weight of scientific evidence in policy and law. *American Journal of Public Health*, 95, S129-S136.
- Leviton, L. C. (2001). External validity. In N. J. Smelser & P. B. Baltes (Eds.), *The international encyclopedia of the social and behavioral sciences* (pp. 5195-5200). Amsterdam, The Netherlands: Elsevier.
- McGlynn, E. A., Asch, S. M., Adams, J., Keesey, J., Hicks, J., DeCristofaro, A., et al. (2003). The quality of health care delivered to adults in the United States. *New England Journal of Medicine*, 348, 2635-2645.
- McLeroy, K. R., Bibeau, D., Steckler, A., & Glanz, K. (1988). An ecological perspective on health promotion programs. *Health Education Quarterly*, 15, 351-377.
- Mohrer, D., Schulz, K. F., Altman, D. G., & Lepage, L. (2001). The CONSORT statement: Revised recommendations for improving the quality of reports. *Journal of the American Medical Association*, 285, 1987-1991.
- National public health performance standards program. (n.d.) Retrieved November 22, 2005, from www.cdc.gov/od/ocphp.nphsp
- Norquist, G. S. (2001). Practice research networks: Promises and pitfalls. *Clinical Psychology: Science and Practice*, 8, 173-175.

- Orleans, C. T. (2000). Promoting the maintenance of health behavior change: Recommendations for the next generation of research and practice. *Health Psychology, 19*, 76-83.
- Ory, M. G., Evashwick, C., Glasgow, R. E., & Sharkey, J. (2005). Pushing the boundaries of evidence-based research: Enhancing the application and sustainability of health promotion programs in diverse populations. In *Enhancing the application and sustainability of health promotion programs*. Edinburgh, UK: Elsevier.
- Pasick, R. J., Hiatt, R. A., & Paskett, E. D. (2004). Lessons learned from community-based cancer screening intervention research. *Cancer, 101*(5), S1146-S1164.
- Pentz, M. A., Jasuja, G. K., Rohrbach, L. A., Sussman, S., & Bardo, M. (in press). Translation in tobacco and drug abuse prevention/cessation research. *Evaluation & the Health Professions*.
- Poland, B. D., & Green, L. W. (2000). *Settings for health promotion: Linking theory and practice*. Thousand Oaks, CA: Sage.
- Prochaska, J. O., Velicer, W. F., Fava, J. L., Rossi, J. S., & Tsoh, J. Y. (2001). Evaluating a population-based recruitment approach and a stage-based expert system intervention for smoking cessation. *Addictive Behaviors, 26*, 583-602.
- Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). New York: Free Press.
- Rohrbach, L., Grana, R., & Valente, T. (in press). Getting innovations to the field. *Evaluation & the Health Professions*.
- Rotheram-Borus, M. J., & Flannery, N. D. (2004). Interventions that are CURRES: Cost-effective, useful, realistic, robust, evolving, and sustainable. In H. Rehmschmidt, M. L. Belfer, & I. Goodyer (Eds.), *Facilitating pathways: Care, treatment, and prevention in child and adolescent health* (pp. 235-244). New York: Springer.
- Schorr, L. B. (1997). *Common purpose: Strengthening families and neighborhoods to rebuild America*. New York: Anchor Books, Doubleday.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental design for generalized causal inference*. Boston: Houghton Mifflin.
- Simons-Morton, B. G., Brink, S. G., Simons-Morton, D. G., McIntyre, R., Chapman, M., Longoria, J., et al. (1989). Ecological approach to the prevention of injuries due to drinking and driving. *Health Education Quarterly, 16*(3), 397-411.
- Simons-Morton, D. G., Parcel, G. S., & O'Hara, N. M. (1988). Implementing organizational changes to promote healthful diet and physical activity at school. *Health Education Quarterly, 15*, 115-130.
- Stange, K. C., Woolf, S. H., & Gjeltema, K. (2002). One minute for prevention: The power of leveraging to fulfill the promise of health behavior counseling. *American Journal of Preventive Medicine, 22*, 320-323.
- Steinberg, E. P., & Luce, B. R. (2005). Evidence based? Caveat emptor! *Health Affairs, 24*(1), 80-92.
- Stokols, D. (1992). Establishing and maintaining healthy environments: Toward a social ecology of health promotion. *American Psychologist, 47*, 6-22.
- Substance Abuse and Mental Health Services Administration. (2005). *SAMSHA model programs: Effective in substance abuse and mental health programs for every community*. Retrieved November 22, 2005, from www.modelprograms.samhsa.gov
- Sussman, S. (2001). *Handbook of program development for health behavior research and practice*. Thousand Oaks, CA: Sage.
- Tinetti, M. E., Bogardus, S. T., Jr., & Agostini, J. V. (2004). Potential pitfalls of disease-specific guidelines for patients with multiple conditions. *New England Journal of Medicine, 351*, 2870-2874.
- Trickett, E. J., & Mitchell, R. M. (1992). An ecological metaphor for research and intervention. In M. S. Gibbs, J. R. Lachenmeyer, & J. S. Sigal (Eds.), *Community psychology and mental health* (pp. 13-28). New York: Gardner Press.

- Truman, B. I., Smight-Akin, C. K., Hinman, A. R., Geldoie, K. M., Brownson, R., & Novick, L. F. (2000). Developing the Guide to Community Preventive Services—Overview and rationale. The Task Force on Community Preventive Services. *American Journal of Preventive Medicine*, 18, 18-26.
- Tunis, S. R., Stryer, D. B., & Clancy, C. M. (2003). Practical clinical trials. Increasing the value of clinical research for decision making in clinical and health policy. *Journal of the American Medical Association*, 290, 1624-1632.
- U.S. Department of Health and Human Services. (2000). *Healthy People 2010: Understanding and improving health*. Washington, DC: Government Printing Office.
- U.S. Preventive Services Task Force. (1989). *Guide to clinical preventive services: An assessment of the effectiveness of 169 interventions*. Baltimore: Williams & Wilkins.
- U.S. Preventive Services Task Force. (1996). *Guide to clinical preventive services* (2nd ed.). Baltimore: Williams & Wilkins.
- Wagner, E. H. (1998). Chronic disease management: What will it take to improve care for chronic illness? *Effective Clinical Practice*, 1, 1-4.
- Walter, L. C., Davidowitz, N. P., Heineken, P. A., & Covinsky, K. E. (2004). Pitfalls of converting practice guidelines into quality measures: Lessons learned from a VA performance measure. *Journal of the American Medical Association*, 291, 2466-2470.
- Ware, J. E., & Kosinski, M. (2001). Interpreting SF-36 summary health measures: A response. *Quality of Life Research*, 10, 405-413.
- Will, J. C., Farris, R. P., Sanders, C. G., Stockmyer, C. K., & Finkelstein, E. A. (2004). Health promotion interventions for disadvantaged women: Overview of the WISEWOMAN projects. *Journal of Women's Health*, 13, 484-502.
- Zaza, S., Lawrence, R. S., Mahan, C. S., Fullilove, M., Fleming, D., Isham, G. J., et al., & Task Force on Community Preventive Services. (2000). Scope and organization of the Guide to Community Preventive Services. *American Journal of Preventive Medicine*, 18, 27-34.