



Research Approaches of Education, Applied Psychology, and Behavioral Science and Their Application to Behavioral Medicine

John P. Allegrante, Christel Hyden,
and Alfgeir L. Kristjansson

Introduction

Behavioral medicine is concerned with the broad and central role of behavior in the prevention and treatment of disease and improvement of quality of life (Fisher, Fitzgibbon, et al., 2011). Because behavioral medicine encompasses interdisciplinary knowledge from across a wide range of theories and best practices in relation to human health and behavioral and social change, it draws extensively on the research methods of education, applied psychology, and behavioral science. These methods are often employed by investigators and program evaluators to make causal inferences about whether

organized intervention efforts designed to change health-related behaviors or the contextual factors that shape behavior result in the desired outcomes. But equally importantly, these methods provide valuable tools for formative research that can inform and guide development of evidence-based behavioral medicine programs, as well as program monitoring for ongoing assessment, improvement, and quality assurance.

In almost any discussion of research methods, scholars and practitioners will invariably get into epistemological debates about the relative value of different methods and different approaches. In the biomedical context and much of behavioral medicine research, historically the randomized controlled trial (RCT) has been seen as the “gold standard” by which novel clinical treatments are evaluated for their effectiveness in achieving a desired health outcome. However, given its standard requirement for a placebo, nonintervention, or alternate-treatment control arm, the RCT model of evaluation is limited in providing evidence about the effectiveness and value of the range of behavioral medicine approaches—approaches that now extend beyond the clinic and into the communities in which people live. As a consequence, research and evaluation efforts in behavioral medicine must go beyond the RCT and utilize different and complementary research strategies, designs, and methods to address health disparities and the myriad complexities of behavior and social factors.

J. P. Allegrante (✉)

Department of Health and Behavior Studies, Teachers College, Columbia University, New York, NY, USA

Department of Sociomedical Sciences, Mailman School of Public Health, Columbia University, New York, NY, USA

e-mail: jpa1@tc.columbia.edu

C. Hyden

Department of Family and Social Medicine, Albert Einstein College of Medicine, Yeshiva University, New York, NY, USA

Harlem Health Promotion Center, Mailman School of Public Health, Columbia University, New York, NY, USA

A. L. Kristjansson

Department of Social and Behavioral Sciences, School of Public Health, West Virginia University, Morgantown, WV, USA

This chapter presents the range of different research strategies, designs, and methods derived from education, applied psychology, and behavioral science that are used in behavioral medicine. The methods discussed in this chapter are often used in conjunction with the methods of RCTs and thus complement the research and evaluation process. Accordingly, we have organized the chapter into several major sections. In the first section, we begin by examining seven key questions that are typically addressed in behavioral medicine and illustrate how the research and program evaluation methods, including qualitative, quantitative, and mixed methods, are utilized in answering each of these questions. In addition to providing a primer on these methods, we include selected examples of how such methods have been applied across a wide and diverse range of problems, practice settings, and populations in behavioral medicine and public health. Although we have organized the sections of the chapter to include descriptions of methods that are principally applied to a specific question or phase of research, it is important to point out that all these methods have overlapping value—whether it is in formative evaluation, process evaluation, or summative evaluation—across the spectrum of the research process, from inception to conclusion. For example, although surveys and focus groups are commonly utilized methods for developing and answering research questions about the problem and the needs to be addressed during the formative phases of research, such methods are often used in the later phases of research and evaluation efforts when results are being interpreted and programs disseminated. In the second section of the chapter, we discuss community engagement in the research process as an integral element of conducting research and evaluation in behavioral medicine, reviewing both the basic concepts and principles of community-based participatory research, including illustrative examples and a discussion of key issues. In the third section, we turn our attention to concepts and issues in the translation of research into practice, a topic of growing interest as a next-generation challenge to those working in both the public health and behavioral medicine communities. Finally, we conclude the chapter with a discussion of some of

the emerging challenges of conducting behavioral medicine research and their implications for the future.

Answering Key Research Questions in Behavioral Medicine

As in most applied behavioral and social research, behavioral medicine is oriented toward: (a) describing and conceptualizing behavioral phenomena, (b) understanding and explaining the causal mechanisms underlying behavioral phenomena, and (c) creating interventional programs designed to facilitate change or some alteration in the behavioral phenomena. Thus, the questions behavioral medicine seeks to address include: What is the problem and what is the need? With what intervention can we best address the problem? How was the intervention delivered? Was the intervention effective, and why or why not? Can the intervention outcomes be replicated? And how can the intervention be improved, scaled, and disseminated?

What Is the Problem and What Is the Need?

Health and human service programs are typically designed to serve people with demonstrated need, facilitate positive human development, and provide prevention and treatment resources to promote or restore health. When developing these programs, *needs assessment* is the critical first step in deciding whether there is indeed a need to be met, identifying the data to be gathered to understand the need, and generating data with which to begin the planning process in determining what types of programs, intervention approaches, or resources should and can be feasibly offered that hold promise to address the need or mitigate the defined problem.

Needs assessment typically examines social, behavioral, and epidemiological profiles of the community. It seeks to identify and address potentially modifiable social problems as well as the agencies, institutions, and programs currently serving the community, or addressing a specific need,

to define and reconcile the discrepancy between “what is” and “what ought to be.” Approaches to answering these questions will vary, but major sources of relevant information in needs assessment include existing archival data about the problem, the experiences and conclusions of experts who know the situation well, and, perhaps most importantly, the perceptions and opinions provided by those directly affected. Once identified, needs can be prioritized and then used as the basis for setting goals and objectives for intervention programs (Isaac & Michael, 1995).

One straightforward method of assessing the needs of a community is simply to ask people about their needs through social surveys, personal interviews, and focus group interviews. At this phase, it is important to estimate the magnitude of need and avoid falling into the trap of proposing and evaluating potential solutions before fully understanding the scope of the problem and the potential objectives a proposed program might be designed to meet. Moreover, the context of need also should be examined in order to ensure efforts are appropriately directed; it is otherwise possible to have an accurate assessment of a community’s need, but fail to assess the cultural, social, or political context in which a program would be implemented or the community’s capacity to continue support and maintenance of the program.

Assessment and Planning Models

Numerous systematic techniques and planning models have been developed to guide needs assessment and the formative stages of intervention planning. While we cannot describe the many techniques and models available, we focus here on two relevant assessment and planning models that have gained popular use and have demonstrated value in a range of health-related and behavioral medicine research.

Delphi Method The Delphi method (or technique) is an iterative process whose purpose is to gather expert opinion and reach consensus of opinion on a defined topic through multiple iterations. It employs a series of questionnaires that are administered by a facilitator in multiple iterative survey rounds that are, in turn, interspersed with feedback and revisions of answers to the

questionnaires in response to the feedback in each round until consensus is reached. Initially developed to improve technological forecasting (Linstone & Turoff, 1975, 2011), the method is especially useful when what is known about a phenomenon of interest is incomplete (Adler & Ziglio, 1996). The Delphi method incorporates components of (a) anonymity of the study’s expert participants and their opinions, (b) multiple rounds of iteration to achieve consensus of opinion, (c) controlled feedback, and (d) the opportunity for quantitative analysis and interpretation of data. Delphi panel studies have been used extensively in health-related research ranging from recent studies designed to clarify concepts of parenting practices around food (Gevers, Kremers, et al., 2014), to the determinants of adolescent coping strategies with cyberbullying (Jacobs, Dehue, et al., 2014), to establishing a framework of behavioral indicators for outcome evaluation of health promotion among individuals with suspected TB patients (Li, Ehiri, et al., 2014).

PRECEDE-PROCEED One of the most prominent and widely utilized models in health planning—PRECEDE-PROCEED—provides a comprehensive planning framework for assessing the health and quality-of-life needs, and for designing, implementing, and evaluating health promotion and other public health programs to meet those needs (Green & Kreuter, 2005). Originally developed to facilitate planning of health education efforts, it has been used extensively and more broadly in health promotion and disease prevention planning, at the national, state and local levels, and globally beyond North America. The model works by guiding program planners and evaluators through an eight-phase series of assessment and analytic steps that result in the formulation of measurable goals and objectives for the program. These are: (a) Phase 1: social assessment and situational analysis; (b) Phase 2: epidemiological assessment; (c) Phase 3: educational and ecological assessment; (d) Phase 4: administrative and policy assessment and intervention alignment; (e) Phase 5: implementation;

(f) Phase 6: process evaluation; (g) Phase 7: impact evaluation; and (h) Phase 8: outcome evaluation. The hallmarks of the model include its flexibility and scalability across health problems, populations, and practice settings; its participatory and iterative nature; and the platform it provides for generating evidence-based best practices. Numerous studies and health program planning efforts have demonstrated its utility. For a bibliography of over 1000 published applications of the model across a wide range of settings, populations, and health problems, see Green (2014).

Quantitative Research Methods

Social surveys provide an opportunity to describe phenomena by examining the responses from a large number of participants and looking for correlations among variables and patterns of cause and effect (McBurney & White, 2010). It is beyond the scope of this chapter to provide more than a brief overview of the survey method. Thus, instead of using illustrations from different levels of inquiry, we will present some general guidelines for the use of survey research in behavioral medicine and some of the considerations the investigator must take into account when planning surveys. There are many excellent texts and other resources on survey methods. Several that may be helpful to those wishing more information are: Aday and Cornelius (2009), *Designing and Conducting Health Surveys*; Dillman, Smyth, and Christian (2008), *Internet, Mail, and Mixed Mode Surveys*; Fowler (2014), *Survey Research Methods*; and Rea and Parker (2014), *Designing and Conducting Survey Research*. In addition, many texts focus on particular aspects of survey research, such as sampling techniques, questionnaire design, question construction, scaling, and data coding and analyses.

Questionnaire Development Survey research relies on the interplay of three key elements in questionnaire development: how a questionnaire is designed, how it is administered, and to whom it is administered. Questionnaire items must be valid, meaning that they must measure what they purport to measure, and they must be reliable,

providing consistently reproducible responses (McBurney & White, 2010). Because questionnaires can be labor-intensive and expensive to develop, frequently researchers will use existing questions or instruments rather than designing their own, allowing them to rely on previous assessment work that has been done to establish the validity and reliability of the questionnaire items and to compare their results with those of previous studies that have employed the same instruments. However, when using existing questions in multicultural populations, items should first be tested for ethnic and racial appropriateness and cultural sensitivity, even if the questions have been used successfully with other population groups (Warnecke, Johnson, et al., 1997). Further, when items are translated to new languages or used in other countries or cultures, additional steps must be taken to ensure the quality of the translation and validity of the items. For example, the SF-36 and SF-12—two widely used measures of health status and quality of life—were developed as part of the RAND Medical Outcomes Study, a multiyear, multisite American study that was originally designed to explain variations in patient outcomes in relation to varying health insurance coverage in the United States (Newhouse, 1982; Ware, Kosinski, & Keller, 1996). These measures have been used extensively in the North American context. But use of these instruments in other countries, cultures, and languages has required additional psychometric testing to establish their clinical validity and to evaluate cross-cultural stability of questionnaire items and scoring algorithms (Bullinger, 1995; Coons, Alabdulmohsin, et al., 1998; Fukuhara, Bito, et al., 1998; Gandek, Ware, et al., 1998; Li, Wang, & Shen, 2003; Ngo-Metzger, Sorkin, et al., 2008; Perneger, Leplège, et al., 1995; Persson, Karlsson, et al. 1998).

When designing new questionnaire items, the investigator will need to consider the purpose of the questionnaire and what they expect to answer or accomplish with the research. For example, is the study seeking to simply describe the presence and characteristics of a phenomenon (e.g., “Who smokes during pregnancy?”) (Schneider, Maul, et al., 2008), to understand why and by what

mechanism the phenomenon might occur (e.g., “What are pregnant women’s knowledge and attitudes towards smoking?”) (Owen & Penna, 2001), or to seek or evaluate potential solutions (e.g., “Is smoking cessation counseling being offered to pregnant women by their healthcare providers?”) (Zapka, Pbert, et al., 2000). In addition, basic principles of questionnaire construction should be followed, including the use of clear, unambiguous items that are valid and reliable, avoidance of bias, logical sequencing, and permitting the data to be coded and analyzed in appropriate and meaningful ways. Steps in questionnaire development often include administering preliminary surveys, called *pilot-test surveys* (where focus groups may be used) to ensure the clarity of questions, determine the correlation between potential items (Zapka, Fletcher, et al., 1997), and conduct psychometric statistical testing to establish construct item validity and reliability or reproducibility (Meadows, Harvey, et al., 2000).

Method of Data Collection Investigators must also choose the method of administration and data collection. These include face-to-face or telephone interviews or self-administered questionnaires (via paper or Internet-based surveys). Each method of data collection will have distinct advantages and disadvantages. For example, face-to-face interviews allow for the development of rapport with a respondent and can be appropriate for respondents with low literacy rates, such as in a study conducted on behalf of the Government of Tanzania in which over 1800 households were interviewed about provision of health services (Abel-Smith & Rawal, 1992). However, face-to-face interviewing is also labor- and resource-intensive, and results depend on both the skill of the interviewer (some of the Tanzanian surveys had to be rejected due to poor interviewing) and on the willingness of the respondent to be honest rather than saying what they think the interviewer wants to hear (i.e., *social desirability bias*).

Self-administered questionnaires can be cost-effective and help avoid social desirability response biases, but can often result in low response rates and missing data. Moreover, there is no concrete

way to know that a respondent understood each question and usually there is little or no opportunity to clarify those questions that may have been misunderstood (McBurney & White, 2010). In using self-administered questionnaires, the researcher also cannot be certain that the intended respondent is the one responding to the questionnaire or if the respondent is acquiring assistance from others. Surveys conducted on the Internet can be highly cost-effective and are often particularly appropriate for sensitive topics since the ability to complete the survey in privacy may result in less social desirability response bias (Cohall, Dini, et al., 2008). Internet-based surveys also include additional benefits, such as controlling for question order that can be administered to respondents based on previous answers. These include branching questions that ask about certain behavior, such as tobacco smoking, and questions thereafter that depend on the answer to the previous question (e.g., if “yes” then answer the next five questions about how much the respondent smokes, preferred brands, etc.; respondents that answered “no” would skip the brand items and be taken directly to the next part of the questionnaire). This kind of control over question sequence can be difficult to accomplish in a self-administered paper-and-pencil survey.

Sampling the Population Investigators must also decide on the kind of sampling design they will use to draw a representative sample from some larger population of interest. This is a critical decision because the generalizability of findings of a study depends heavily on the extent to which a sample may be truly representative of the population. The first set of these designs is *nonprobability samples* and include *purposive samples*. *Purposive samples* are designed to identify potential respondents for some particular purpose. For example, research intending to describe the experiences of Canadian adults with osteoarthritis (Gignac, Davis, et al., 2006) or gay Scottish men with HIV (Flowers, Duncan, & Knussen, 2003) would necessarily seek out those specific populations. Similarly, a study of nursing students’ perceptions and health beliefs may focus on the students at a single university (Denny-Smith, Bairan, & Page, 2006), research on patient or pro-

vider knowledge and experiences may utilize only individuals at specific hospitals, clinics, or pharmacies (Bakken, Holzemer, et al., 2000; Johnson, Nowatzi, & Coons, 1996; Parker, Baker, et al., 1995; Pole, Einarson, et al., 2000; Secginli & Nahcivan, 2006), and surveys of specific professions may be distributed to potential respondents by using membership directories as sampling frames (Helft, Hlubocky, & Daugherty, 2003; Kenny, Smith, et al., 1993; Story, Neumark-Sztainer, et al., 2002) or at trade shows and professional meetings (Korelitz, Fernandez, et al., 1993). Other examples of nonprobability samples that have been identified include *Quota samples*, which seek to identify a certain number of respondents in the sample; *chunk samples* (sometimes referred to as *convenience samples*), which study a group of respondents who happen to be available; and *snowball samples*, which is “chain sampling” that starts with a single or small group of initial respondents (often useful in hard-to-reach populations such as drug users or sex workers) who then identify potential other similar respondents (Aday & Cornelius, 2009). In each case, external validity, i.e., the potential to generalize to a larger known population will vary depending on the sampling design.

The second type of sampling design, which is more powerful, is *probability samples*. In using probability samples, researchers generally obtain respondents in a systematic manner such that the probability that any given individual within a defined universe (sometimes referred to as the *sampling frame*) of potential respondents representing a *population* will have an equal chance of appearing in the sample. *Simple random samples*, *systematic samples*, *stratified random samples*, and *cluster samples* all provide different probability-based methods for selecting appropriate population-based samples from the group of interest (Aday & Cornelius, 2009). National studies on knowledge, attitudes, behavior (Galuska, Will, et al., 1999; Glasgow, Eakin, et al., 2001; Knuth, Malta, et al., 2011; Lantz, House, et al., 1998), disease prevalence (Burney, Luczynska, et al., 1994; Tsugane & Sobue, 2001; Yang, Lu, et al., 2010), and their correlations typically use national probability samples, which allow investigators to

generalize to the larger population within a statistically known margin of error (McBurney & White, 2010). In this way, investigators may calculate the probability that any one sample is not completely representative of the population from which it has been drawn; while sampling error cannot be eliminated, the extent of the error will be influenced by the sampling techniques chosen (Kelley, Clark, et al., 2003).

Collecting survey data, however, can be expensive and time-consuming. In some cases, the existing specialized and publicly accessible national data sets, which constitute a source of household and individual survey data for secondary analyses, can permit investigators to answer important questions without the time and expense of collecting new data. In the United States, the National Health Interview Survey (NHIS) (CDC, 2015a) includes US Census data to track health status, health-care access, and progress toward achieving national health objectives among household respondents; the Behavioral Risk Factor Surveillance System (BRFSS) (CDC, 2015b) comprises a monthly cross-sectional telephone survey that state health departments conduct with a standardized questionnaire to collect prevalence data from the adult US population on risk behaviors and preventive health practices that affect their health status; and the Health Information National Trends Survey (HINTS) (NCI, 2015) routinely collects nationally representative data about the American public’s use of cancer-related information and related topics. Some surveys, like the National Health and Nutrition Examination Survey (CDC, 2015c), combine individual interviews with physical examinations of a subsample of respondents to assess the health and nutritional status of adults and children. Such sources of national data have supported studies on a wide range of health topics, including characteristics related to participation in a smoking cessation trial (Graham, Papandonatos, et al., 2008); the influence of lifestyle on inflammation in men and women with type 2 diabetes (Jarvandi, Davidson, et al., 2012); and the relationship between physical activity and general mental health (Kim, Park, et al., 2012).

Finally, sampling designs can be mixed. Some studies, for example, mix sampling techniques by

randomly sampling from within a sampling frame such as a professional directory or other listing of potential respondents (Stanwood, Garrett, & Konrad, 2002; White, Speechley, et al., 1995) or from purposively chosen study sites (Kurth, Kamtsiuris, et al., 2008). A study of the health of homeless children and housed, low-income children in Los Angeles utilized such mixed sampling techniques. The study employed a three-stage sampling strategy: (1) a purposive sample of shelters, (2) a systematic sample of families in shelters, and (3) a random sample of one child in each family (Wood, Valdez, et al., 1990). On a broader geographical scale, an investigation of suboptimal utilization of public health facilities in Afghanistan began with a sampling frame of six provinces, within which two districts were selected using mixed sampling technique representing urban and rural populations. Two community health centers (CHCs) were selected within each district and out of the two, one from the center of the district was identified and another from the broader geographical catchment area of the district. At the level of CHC, two villages were then selected: the village in which that CHC itself was situated and another study village from the CHC catchment. In this way, at the level of selection of study sites, priority was given to select a fair percentage of respondents from rural as well as urban areas. Using this method, the investigators gave priority to selecting a fair percentage of respondents from rural as well as urban areas and resulted in a total of 48 villages and 24 health facilities from 12 districts in 6 provinces (Singh, Sharma, et al., 2012). Similarly, an evaluation of the quality of public health services in India had no sampling frame from which to draw respondents. Based on the literature, the investigators chose a sample size of 500 respondents to be drawn from the state of Uttar Pradesh, which was divided for sampling purposes into three geographic regions: eastern, central, and western regions. The sample size of 500 was distributed to these three regions in proportion to the rural population of the respective region, and two districts representing each region were selected randomly and the number of respondents selected from each district was proportional to the rural population of the respective

districts. Finally, inclusion criteria required that the respondent should have utilized services at the public health center in the previous 6 months, and initial identification of sampling units was accomplished by seeking referrals from village leaders and the medical staff at the health centers, with subsequent respondents being identified through snowball sampling (Narang, 2011).

Qualitative Research Methods

In comparison to the descriptive and correlational nature of the data obtained in surveys, qualitative research is usually exploratory and seeks to use inductive (starting with observations and developing hypotheses) rather than deductive (starting with extant hypotheses and testing them with observations) approaches to generate novel insights. Curry, Nembhard, and Bradley (2009) have noted that such methods are best utilized when (a) investigating complex phenomena that are difficult to measure quantitatively, (b) generating data necessary for a comprehensive understanding of the problem, (c) gaining insights into potential casual mechanisms, (d) developing sound quantitative measurement processes or instruments, and (e) studying special populations.

Qualitative research differs from quantitative research in that rather than counting occurrences, exploring correlations among variables of interest, and statistically testing hypotheses, qualitative research seeks to describe the complexity and range of occurrences or phenomena and provide a rich basis for generating hypotheses or gaining deeper insights into statistically demonstrated relationship among the variables of interest. Moreover, while quantitative research typically generates numeric data using standardized processes and instruments with predetermined response categories, qualitative research allows for the use of open-ended questions, discussions, and observations. These guided discussions also allow the respondent to identify, describe, or elaborate on concepts and concerns that may not have been previously anticipated by the investigators and not captured with the more closed-format questions of those found in surveys (Curry, Nembhard, & Bradley, 2009).

The two primary methods of qualitative data collection are in-depth interviews and focus group

interviews. In-depth interviews allow for the exploration of individual experiences in great detail and can be particularly valuable for sensitive topics since the method maximizes privacy while also allowing the investigator to build rapport to increase candor. Focus groups are equally well suited for explorations of perceptions and traditions of social groups and understanding social processes, as the group interaction dynamic can serve as a catalyst to generate unique insights into understanding (Kreuger & Casey, 2015; Mermelstein, 1999). Research suggests that group discussions can also elicit more critical comments than interviews, with the synergy of the group allowing each participant to reinforce one another's vented feelings (Robinson, 1999). In this way, focus groups may be especially effective at facilitating comfort among socially marginalized or disempowered populations who might otherwise feel reluctant to give negative feedback or who may feel that any problems result from their own shortcomings (Curry, Nembhard, & Bradley, 2009; O'Brien, 1993; Robinson, 1999).

Examples of research in which such methods have been applied include studies designed to illuminate the health beliefs and folk understanding regarding diabetes among British Bangladeshis (Greenhalgh, Helman, & Chowdhury, 1998) and formative research processes focused on intervention development for hard-to-reach population groups by the AIDS Community Demonstration Project (Higgins, Oreilly, et al., 1996). Because of the increasing importance of such methods in health-related research, efforts to formulate and define standards for reporting qualitative research have been undertaken in recent years. For example, the 21-item standards for reporting qualitative research (SRQR) is designed to improve the transparency of qualitative research by providing clear standards for reporting of study methods and findings (O'Brien, Harris, et al., 2014).

As with quantitative methods, investigators using qualitative research methods must carefully define the target group that will be the most beneficial, using systematic scientific methods to develop the sample (Curry, Nembhard, & Bradley, 2009; Robinson, 1999). However, in contrast to quantitative sampling techniques that rely on statistical

probability theory, the logic and power of the purposive sampling used in qualitative research lie primarily in the high quality of information obtained per sampling unit. Adequacy of the sample size is relative, a matter of ensuring that the sample is neither too small to support claims of informational redundancy or saturation, nor too large to allow the deep, case-oriented analysis that is the hallmark of so much qualitative work (Sandelowski, 1995). Thus, generally, the aim is to identify participants who are "information rich," have certain characteristics, possess detailed knowledge, or have relevant experience; to study their responses intensively; and to continue data collection until the point of theoretical saturation, i.e., when no new concepts emerge (Curry, Nembhard, & Bradley, 2009). Although it is not possible to define the number of participants in advance, a range of 20–30 individual interviews or 4–6 focus groups with 6–10 participants each is often adequate to achieve saturation (Morgan, 1996; Patton, 2002). However, studies involving more than one target population, more heterogeneous groups, or both, often require more episodes of data collection in order to ensure inclusion of multiple viewpoints. Examples include the AIDS Media Resource Project, which conducted 52 different focus groups with 351 participants (Kitzinger, 1994), and a multisite investigation into ethnic and gender differences in youth smoking which included 178 focus groups conducted in 11 states with 1175 participants (Mermelstein, 1999).

Finally, as with quantitative methods, investigators using qualitative methods should strive to maintain that same scientific rigor that typically characterizes quantitative research. This means aiming to reduce problems such as researcher bias, lack of reproducibility, or limited generalizability.

Mixed Methods

Increasingly, and in part due to a recognition of the complexity of the problems on which behavioral medicine focuses, current evaluation research practice is now emphasizing a blending of the two approaches in which qualitative findings add interpretive richness to the more objective findings of quantitative research. Such *mixed-method* approaches in health-related research have emerged in recent years (Chatterji, 2004;

Chatterji, Green, & Kumanyika, 2001; Clark, 2010; Peterson, Czajkowski, et al., 2013). The inherent strengths of quantitative and qualitative research approaches complement each other, and combining both methods not only reflects the complex nature of the problems facing behavioral medicine and public health, but can also improve the quality and scientific power of the data derived from the investigation of complicated health problems (Creswell, Klassen, et al., 2011). Examples of mixed-methods research include studies of vaccine reminders (Anderson, Sebaldt, et al., 2008), patient safety (Benning, Ghaleb, et al., 2011), HIV and sexually transmitted disease (STD) prevention (Pinto & McKay, 2006; Shain, Piper, et al., 1999), and smoking cessation (de Vries, Weijts, et al., 1992). Combining these two approaches can result in a synergistic effect where the outcome of the two together is greater than the effects of either approach used separately (de Vries, Weijts, et al., 1992).

Reports of qualitative findings enriching understanding of quantitative observations have also become more numerous in the literature in recent years. For example, in the Feeding Young Children Study, a randomized controlled trial of a bottle-weaning intervention among low-income families, the initial research questions at Women, Infants, and Children (WIC) nutritional clinics were formed via nutritionists' observation of 4- and 5-year-old children drinking from baby bottles. A pilot quantitative study confirmed that mothers did typically provide baby bottles to children well past the recommended weaning age (Bonuck & Kahn, 2002), but it was the subsequent focus group discussions that revealed the mothers were typically following feeding advice from the child's maternal or paternal grandmother. Moreover, while the mothers were open to changing behavior and learning new skills, they had concerns about implementing changes that were counter to the grandmothers' opinions and experiences. The qualitative findings put the quantitative child-feeding data into the larger context of family dynamics, and the subsequent intervention not only addressed the mothers' knowledge of feeding behaviors but also provided support and materials to help mothers to

broach the topic at home with family members (Hyden, Kahn, & Bonuck, 2013). In this case, the insights gleaned from the qualitative research shed new light on the quantitative results and clarified the information needs of the target audience.

Mixed methods can also provide researchers with additional tools to validate the outcomes of studies. Sometimes referred to as *triangulation of methods*, if the results from each method suggest the same conclusion, then confidence in the results is strengthened (Steckler, McLeroy, et al., 1992). One evaluation design solution that encompasses several phases of evaluation research using mixed methods is the extended term mixed method (ETMM) approach. ETMM designs are long-term research plans following life spans of individual programs or policy initiatives by employing descriptive research methods in the early stages of program adoption and implementation followed by experimental designs at a subsequent stage. ETMM designs deliberately study and document environmental variables as a component of the research plan, allowing for explanations of causality based on both empirical and substantive knowledge gained on the program and its setting. This use of a variety of research methods at multiple points of a program or policy's life span can improve the quality of evidence and strengthen interpretations of causality by helping to shed light on a multitude of context, process, and input indicators. The investigator can also select the key variables and interactions to use as statistical or procedural controls to empirically test process–outcome links (Chatterji, 2004). For example, an evaluation of after-school supplemental education utilizing ETMM involved a year-long study integrating a matched-groups design with classroom observations and surveys. The research began with a 14-week formative phase conducted at the beginning of the semester to explore the program and its environment in depth, with the goal of providing feedback to developers, program personnel, and school staff in order to stabilize treatment delivery and improve fidelity. This “before” phase gathered process data using classroom observations and teacher surveys, and yielded evidence of the extent to which the observed program processes, inputs,

and outcomes were consistent with the program's underlying theory and philosophy. In the last 16 weeks of the program, data collection continued with classroom observations and surveys to document changes on program inputs and processes over time in matched classrooms by grade—the summative or “after” phase. In this design, the findings of the formative phase were used to tighten the data-gathering and analytic design of the summative phase, and qualitative classroom observations triangulated the quantitative analysis of teacher surveys and student outcome measures (Chatterji, Kwon, & Sng, 2006).

In summary, the use of quantitative, qualitative, and mixed research methods in behavioral medicine research is critical to gaining an understanding of the problem and needs of the population. The use of such methods in systematic needs assessment enables investigators and program planners to gain an understanding of the scope and extent of the problem, identify potentially feasible approaches to addressing the problem and perceived needs, and set the stage for the formulation and development of intervention programs. Mixed methods are important in gaining a better picture of the impacts and outcomes of intervention programs, often providing insights into the barriers and enablers to intervention success.

Table 6.1 provides a summary of the three research approaches and methods for needs assessment discussed above, including selected characteristics by which the strengths and weaknesses of each method can be assessed.

With What Intervention Can We Best Address the Problem?

Once a problem has been identified, the question of what intervention can best address the problem then becomes the major focus of program developers. Answering this question relies heavily on both formative and process evaluation methods (Dehar, Casswell, & Duignan, 1993). Program and intervention development is typically shaped during a *formative evaluation* phase of research. This phase usually seeks to address questions about program design, process, and

outcomes, and identify elements of the intervention approach that will be necessary to change what is known from a needs assessment about the potentially modifiable causal mechanisms of the behavior or circumstances that are the focus of change. During this process, behavioral researchers not only need to understand the incidence and prevalence of a problem but must also gather data that can answer several key questions about how that problem can be optimally addressed.

Formative research typically cannot be conducted from a distance and must include input from the community of interest, as well as from those who have access to and knowledge about the intended audience, in order to create programs that are acceptable, unique, and effective (Posavac, 2011). Formative evaluation research activities can include having members of the intended audience evaluate materials for clarity and effectiveness, conducting surveys or interviews with potential partners or participants to inform the direction and content of program activities, and pretesting recruitment strategies, data collection methods, and pilot intervention delivery with small groups representative of the larger intended audience (Dehar, Casswell, & Duignan, 1993). For example, Peterson, Link, et al. (2014) conducted a three-step approach to developing and evaluating a novel coronary artery disease self-management educational workbook to be used in a novel intervention being tested in a randomized controlled trial (Peterson, Charlson, et al., 2012). First, the investigators conducted interviews using grounded theory methods with a diverse cohort of patients to identify needs and perceptions. Second, they then incorporated the themes that emerged from the qualitative interviews into the design of the workbook content. Finally, they evaluated study participants' use of and experience with the workbook at the end of the 12-month study period, demonstrating that the focus on practical health information, behavior-specific self-efficacy, and how healthy behaviors decrease risk was highly relevant to achieving study outcomes.

In another example, within the Trial of Activity for Adolescent Girls (TAAG), a randomized, multicenter field trial to reduce the decline in physical activity in adolescent girls, each field center

Table 6.1 Summary of research approaches and methods and selected design characteristics

Characteristic	Quantitative methods	Qualitative methods	Mixed methods
Scientific method	Deductive or “top-down”: the researcher uses data to test or verify hypotheses and theories	Inductive or “bottom-up”: the researcher uses data collected in the field to generate new hypotheses and grounded theory	Deductive and inductive
Objectives	Description, explanation, prediction	Description, explanation, discovery	Multiple objectives
Nature of research	Objective; attempt to study behavior under controlled conditions	Subjective, personal, socially constructed; attempt to study behavior in natural environments and understand the context within which behavior occurs	Pragmatic; attempt to study behavior in multiple contexts or conditions
Nature of data and sources	Variables from surveys and experiments	Words, images, categories from documents, observations, interviews, ethnography, field notes	Mixture of variables, words, and images
Lines of inquiry	Closed-ended questions with predetermined approaches	Open-ended questions with a focus toward emerging approaches	Both open- and closed-ended questions focused on emerging and predetermined approaches
Data analysis	Correlation and other inferential statistical relationships	Description and identification of themes and patterns	Quantitative and quantitative presentation of multiple perspectives
Data results	Researcher framed; generalizable findings with a general understanding of respondent’s viewpoint	Respondent framed; particularistic findings with an in-depth understanding of respondent’s viewpoint	Corroborated findings that may be generalizable
Data reporting requirements	Statistical report including correlations, comparisons of means, and statistically significant findings	Narrative report including contextual description, categories, themes, and supporting respondent quotes	Statistical findings with in-depth narrative description and identification of overall themes
Strengths	Uses standards for validity and reliability; relates variables to hypotheses; utilizes unbiased approaches; provides uniform information; can draw correlations between variables and examine changes in variables over time; generalizable and useful for making predictions, particularly when replicated; useful for studying large groups of people	Focuses on concepts, contexts, and settings; allows for exploration and explanation of personal values and experiences using participants’ categories of meaning; can allow for more collaboration with participants; useful for understanding complex phenomena; responsive to local situations and can include in-depth analyses of contexts and environments; flexible and adaptable	Allows for integration of different data at different stages of inquiry; provides context to quantitative findings to supplement findings by allowing individual cases to illustrate larger findings; can add insights and understanding that might be missed with only a single method

(continued)

Table 6.1 (continued)

Characteristic	Quantitative methods	Qualitative methods	Mixed methods
Weaknesses	Susceptible to respondent bias; focus on hypothesis testing may obscure observation of other phenomena (i.e., confirmation bias); researchers' categories may not reflect respondents' contexts; difficult for use with complex or conceptual issues; results may not translate well to specific situations, contexts, or individuals; requires adequate sample size for statistical power	Data collection and analyses can be expensive and/or time-consuming; analyses are subjective and can be more easily influenced by biases and idiosyncrasies; results are not generalizable; difficult to use to make predictions; lack of stringency may not meet acceptability or credibility criteria for some the scientists, community leaders, etc.	Can be difficult for a single researcher to carry out both qualitative and quantitative research, both in terms of workload and skill; can be expensive and/or time-consuming; methodological challenges, e.g., paradigm mixing, interpreting conflicting results

Note. Adapted from Johnson and Christensen (2004)

worked with schools and communities which differed appreciably in geography and ethnic/racial and cultural backgrounds. The multiphase, mixed-methods TAAG formative evaluation research protocol was developed to address these complexities while understanding how to maximize acceptability by schools, parents, and students to enhance potential for program sustainability. The first phase included (a) school surveys to determine physical education (PE) and health education requisites, teaching strategies, physical activity facilities, and after-school programs; (b) surveys of community agencies to identify resources, communication strategies, and the role of staff; (c) a parent survey to determine the parents' and girls' physical activities, access to resources, physical activity barriers, and preferred methods of learning about programs; (d) a girls' activity checklist to determine prevalent and favorite physical activities; (e) in-depth interviews with girls to determine their favorite activities, barriers to being active, social and environmental contexts, and attitudes about PE; and (f) focus groups with boys to understand their perceptions of girls being active. The second phase included focus groups with girls and interviews with PE instructors to refine the development of intervention materials, define meaningful segments for tailoring intervention messages, explore potential channels for delivering intervention messages, and understand the resources and

constraints of target school PE departments. By including multiple respondents and data collection methods, the TAAG approach allowed a greater understanding of physical activity in adolescent girls from a variety of perspectives, including teachers, parents, community agencies, boys, and the girls themselves. Thus, the structured modes of data gathering during the formative evaluation phase produced important information that might otherwise not have emerged had representatives of the intended population not been consulted (Gittelsohn, Steckler, et al., 2006).

Similarly, development of a nutrition education program for use in Red Cross chapters throughout the United States went through a multistage process that began with analysis of the program content for technical accuracy and sequencing of materials before having potential instructors and course participants critically examine the program. Formative evaluation research activities then moved into formal pretests of the materials, teaching strategies, and survey instruments, first at 6 sites, then at 10, and finally with a national field test at 51 sites. The program was modified after each stage of the process based on the data obtained before a full-scale implementation was launched on the national level (Dehar, Casswell, & Duignan, 1993; Edwards, 1987). This illustrates two important points: First, there is not always a clear boundary between formative evaluation

research methods and the summative methods that are used to assess effectiveness; second, researchers and practitioners can frequently get bogged down in concerns about following rigid rules and making distinctions. This suggests that understanding how underlying processes that are at play may influence the objects of interest in evaluation should be the focus, not allegiance to any one particular method or another.

Other formative evaluation research activities may guide the adaptation or modification of existing programs and resources for new audiences. For example, a behavioral intervention to prevent STDs among minority women was based in part on ethnographic data collected through focus groups; interviews; observations on life and lifestyles, values and beliefs, sexual behavior, knowledge, and risk taking; strategies to motivate behavioral change; and the logistics of implementing a potential intervention. These findings were then integrated into a pre-existing AIDS reduction model to create a new culture- and sex-specific small group intervention (Shain, Piper, et al., 1999).

When utilized most effectively, formative evaluation research is an *ongoing* process that is *integrated* into the development and implementation of a research project, providing assessment information within a feedback loop that identifies the strengths and weaknesses of the project and its intervention approaches as it evolves, informs modifications of measurement instruments, and shapes the evaluative research design and the intervention program (Evans, Raines, & Owen, 1989). Thus, well-designed formative evaluation research will inform monitoring of intervention program delivery as well as the outcomes and process evaluation activities that are used to assess whether efforts and resources are directed as needed and planned and are of sufficient quality and intensity to achieve desired goals for change.

How Was the Intervention Delivered?

Once a program or intervention has been developed and is being implemented, the focus of attention turns to *process evaluation*. This includes evaluating the extent to which the program is

being delivered as it was designed (often referred to as *treatment* or *program fidelity*) and the degree to which the program is functioning as designed and achieving the expected goals and objectives. McGraw, McKinlay, et al. (1989) have described five functions common in process evaluation. These include identifying the: (a) extent to which a program reaches the target population; (b) program dose, i.e., frequency of delivery and/or participation in program activities; (c) organizational context or variability within which the program is being implemented; (d) extent to which programs are implemented in line with program goals; and (e) cost of program implementation. Thus process evaluation validates the assumptions made during the program planning stages to ensure that the needs of the intended population as they were identified in the needs assessment and formative evaluation phase are being met and that program activities are being implemented as designed. Seeking discrepancies between the program delivery plan and the reality of implementation allows researchers either to continue implementing the program with fidelity to its original design or modify it appropriately to adapt to the realities or unanticipated barriers encountered during implementation. This is critical because in many studies oriented toward developing and evaluating effective interventions, the planned intervention protocol may need to be adapted in response to the emerging data obtained from process evaluation. Process evaluation can also include a comparison of program plans to actual operation, identification of specific program components that appear to most influence outcomes, and an analysis of the internal dynamics of a program to understand its strengths and weaknesses and the changes in these that occur over time (Dehar, Casswell, & Duignan, 1993). The key to successful process evaluation is making ongoing, careful observations during implementation (especially in the early phase), according critical consideration to the evidentiary weight of these observations, and introducing systematic adjustments in the program or its implementation in response.

It is a mistake, however, to plan evaluation research that focuses solely on the long-term, summative evaluation of program outcomes, and

ignore its process, performance, and immediate impact. Thus, two important functions of process evaluation are: (a) assisting in the interpretation of outcomes, and (b) informing future efforts in similar areas. If a program fails to show impact and lacks process evaluations, it will be impossible to know whether the lack of impact is a reflection of failure of theory and program design, failure to implement the program as originally specified, or failure of measurement to detect program impact or effectiveness (Weiss, 1972). If, however, a program is shown to be successful, detailed information about what it consisted of and how it was implemented will be critical for replication and dissemination. Perhaps most common is a combination of the two scenarios: If a program has mixed success achieving its goals, detailed information about program operations will be necessary to identify and adopt in future programs only those features that were successful (Dehar, Casswell, & Duignan, 1993). In short, while summative outcome measures may illuminate which programs perform well and which interventions were associated with a given outcome, it is equally important to know what “key ingredients” (or components) determine a program’s success and how those who manage and participate in the program think and behave (Lindsay, 2002).

To illustrate, the “Gimme 5: A Fresh Nutrition Concept for Students” program was a 4-year intervention targeting increased fruit and vegetable consumption by high school students that utilized multiple components: (a) a school-based media campaign, (b) classroom workshops, (c) school meal modification, and (d) parental involvement. For each of these four intervention components, process evaluation strategies were developed to assess program dose, penetration, and utilization, as well as external competing factors. Data collection methods included questionnaires, classroom observations, measurements of student attendance, and assessment of school menus, food offerings, and food use. The process evaluation results not only demonstrated that the intervention was implemented as planned, but also showed how variability in program dose, penetration, and utilization of a multicomponent intervention can influence the outcomes (Nicklas & O’Neil, 2000).

Similarly, a comparison of two variations of a nurse-led psychoeducational intervention to assist oncology outpatients to manage their pain integrated process evaluations via a qualitative study embedded within a RCT of patient outcomes. Using audiotapes of the intervention sessions along with nurse and patient notes to describe the issues, strategies, and interactions experienced during the intervention, the researchers were able to evaluate not only the outcomes of the intervention, but the process of delivering it (Schumacher, Koresawa et al., 2005). In another example, a study of restrictive smoking policies used surveys of employees and supervisors administered before and after the date the policy became effective as primary outcomes measurements. However, qualitative data, including written comments on surveys, focus groups, and structured interviews, were used to elucidate the findings and identify themes and program characteristics which appeared to have the strongest influence on outcomes of the policy (Gottlieb, Lovato, et al., 1992).

Finally, process evaluation research bridges the gap between the intervention design and its impact and outcomes, providing a more comprehensive and well-rounded approach to program evaluation. To illustrate, the SPARK program, a controlled field study of a multicomponent elementary school program to promote physical activity, included a weekly classroom-based self-management program designed to teach behavior change skills such as goal setting and self-instruction to help children generalize physical activity outside of school. An evaluation of curriculum implementation and association between process and outcome was conducted using direct observation of lessons, subjective ratings by teachers and parents, and participation records of students. Investigators found that teachers viewed the self-management curriculum less positively than the physical education curriculum, and teachers were observed implementing the self-management curriculum at an average rate of 65%. Both of these findings may have contributed to the limited outcome effects of the self-management program. The process evaluation thus allowed program coordinators to identify barriers to full program implementation, which

could then be used to inform future iterations of the intervention with the goal of improving curriculum implementation by teachers (Marcoux, Sallis, et al., 1999).

Was the Intervention Effective, and Why or Why Not?

The practice of program evaluation research incorporates the systematic collection of data about program characteristics, activities, impact, and outcomes to improve effectiveness, and reduces uncertainty in making decisions regarding what the program does and what it affects (Patton, 1987). If process evaluation has demonstrated how a program has been implemented, an assessment of the impact and outcomes becomes the next step in program evaluation research. *Summative evaluation* is concerned with effects of intervention programming that is both proximal (*impact*) and distal (*outcome*). Moreover, summative evaluation is complicated by the challenges of attributing the causes of behavioral change to intervention (especially in the absence of a control condition or comparison group) and the differing opinions stakeholders will have about what constitutes a successful outcome or how long that outcome must be sustained for the program to be considered a success (Posavac, 2011). In the following sections, we take up concepts of impact and outcome evaluations and provide examples of their application across several settings.

Impact

Much of summative evaluations research focuses strictly on reported changes in attitudes, behavior, or immediate clinical outcomes that are proximal to program implementation.¹ For example, in an STD and HIV intervention program that priori-

tized female sex workers in China, a women's health clinic was set up near various sites of participants' work (e.g., karaoke bars, massage parlors, and dance halls). Cross-sectional surveys at baseline and postintervention revealed that the rate of condom use with the most recent three clients increased from 55% at baseline to 68% 12 months later, and the prevalence of gonorrhea and chlamydia fell from 26% and 41%, respectively, to 4% and 26%. These results were used to develop national guidelines on sex worker interventions for nationwide replication (Rou, Wu, et al., 2007).

Investigators evaluating an AIDS prevention program for American sex workers found that a mixed-methods approach to evaluation was the most appropriate fit for the impact evaluation of their program. Field staff indigenous to the neighborhood and population was utilized to readily gain access to the community of sex workers; research methods included open-ended interviews with participants and ethnographic field notes, as well as epidemiological questionnaires. This approach allowed respondents to share, in their own words, their feelings about risks for AIDS, which provided primary findings about higher condom-use behaviors with clients versus lower use with husbands or boyfriends. The approach not only addressed research problems endemic to street-based populations but also ultimately provided a more comprehensive assessment of the program's impact than either method could have provided alone (Dorfman, Derish, & Cohen, 1992).

In another example, when Brazilian researchers sought to test the effectiveness of a program designed to improve child growth by training health workers in nutrition counseling, they randomized children to health facilities with trained workers and compared them to those attending facilities with standard care (Santos, Victora, et al., 2001). The research demonstrated that children receiving the intervention had statistically significant weight gain compared to the control group. However, the impact of behavioral programs often depends on factors outside the health system. In this case the researchers had to demonstrate at least six levels of impact, including that it was possible to train many workers in the intervention, that

¹When referring to the effects of intervention, the terms *impact* and *outcome* are used somewhat differently in the behavioral and social sciences than in clinical medicine. In public health, proximal effects are often referred to as impacts, whereas more distal effects are called outcomes. In clinical medicine, however, proximal effects are outcomes and the longer-term effects of treatments are thought of as the impact.

mothers were receptive to and understood the messages they received, and that not only the mothers changed their child-feeding behavior but that the children ate the more nutritious food (Victora, Habicht, & Bryce, 2004).

Other impacts of such evaluative research will include measurements of program information reach to the intended audience in an effort to assess relationships between awareness and behavior. For example, the PSI/PMSC Horizon Jeunes was a youth-targeted social marketing program for improving adolescent reproductive health in urban Cameroon through peer education, youth clubs, mass media promotion, and other behavior change communications. Using preintervention and postintervention surveys at an intervention and comparison site, the investigators found that after about 1 year of intervention, knowledge of the program was nearly universal, and the majority of youth had direct contact with the program. Exposure to the intervention had a significant effect on several proximal determinants of preventive behavior, including awareness of sexual risks, knowledge of birth control methods, and discussion of sexuality and contraceptives, as well as an increase in the proportion of young women who reported using oral contraceptives and condoms for birth control (Van Rossem & Meekers, 2000).

Outcomes

Compared to immediate impacts on variables of interest such as attitudes and behavior, the effort to evaluate the longer-term effects of intervention programs on more distal health outcomes, such as health status or quality of life, can be significantly more challenging, as illustrated in the following examples of outcomes evaluation from three different settings.

Community-Based Cardiovascular Risk Reduction Some of the best examples of evaluation of long-term health effects have emerged from the several historic studies of NIH-funded community-based cardiovascular disease (CVD) risk reduction experiments that have been conducted in North America and in Europe. In North America, studies conducted at Minnesota (Luepker, Murray, et al., 1994), Pawtucket

(Carleton, Lasater, et al., 1995), and Stanford (Farquhar, Fortmann, et al., 1990) all utilized a common theoretical basis—social learning theory—in designing a multiple-component intervention approach that included mass media and social marketing, community organization, and direct education of health professionals. All the projects were evaluated using quasiexperimental designs, with intervention cities and comparison cities. In the Stanford Five-City Project, comprehensive community health education whose aims were to reduce community CVD risk factors was conducted in several cities in northern California from 1979 through 1992. The intervention addressed multiple cardiovascular risk factors and was delivered to all residents in two treatment communities from 1980 to 1986, using multiple educational methods. To evaluate outcomes, potentially fatal and nonfatal myocardial infarction and stroke events were identified from death certificates and hospital records abstracted from hospital charts, coroner records, physicians, and next of kin. Over the full 14 years of the study, the combined-event rate declined about 3% per year in all five cities. However, during the first 7-year period no significant trends were found in any of the cities; it was only in the final 7-year period that significant downward trends were found in all except one city. The change in trends between periods was in the hypothesized direction but not statistically significantly greater in the treatment cities than comparison cities. The researchers speculated that some other influence (e.g., secular trends) accounted for the observed change in all the study communities (Fortmann & Varady, 2000).

Similarly, the North Karelia Project, a study of CVD prevention in Finland, encouraged community action that enabled local community health coalitions and public health departments to do whatever they could imagine would make sense to bring about community and individual changes in health-related behavior (Puska, Nissinen, et al., 1985; Puska, Vartiainen, et al., 1998). The program was developed by Pekka Puska and his colleagues in the Department of Epidemiology of the National Public Health Institute with field offices at the level of county departments of

health and local advisory boards in North Karelia. Community organization in North Karelia included collaboration with existing official agencies and voluntary health organizations so that the new project activities in CVD prevention could be integrated with ongoing, formal public health activities (Puska, Nissinen, et al., 1985). Like many North American CVD risk reduction projects, the North Karelia project set a strong example for the use of multiple channels and intervention approaches, from mass media to cooperation with agricultural and food merchandising groups, for example, to improve the availability of healthy foods such as low-fat milk and other products (Puska, Nissinen, et al., 1985). Mass media interventions included the production of health education materials and messages that were disseminated through local newspapers and community organizations and campaigns. Training activities included not only doctors and nurses but also social workers, representatives of voluntary health organizations, and informal opinion leaders. Training was organized through county-level or other local organizations. Attention to the health system included reorganizing treatment for hypertension and care following myocardial infarction (MI). This included training and development of treatment guidelines. Cooperation with other local organizations included not only the voluntary health agencies but also the critical food industry (e.g., including dairies and sausage factories) and grocery stores (Puska, Nissinen, et al., 1985). Two characteristics appear critical in the North Karelia community organization: (a) the variety of activities and channels included and (b) the attention in all areas to implementation through and in collaboration with local organizations. In comparison to other parts of Finland, the North Karelia campaign led to impressive reductions both in CVD risk factors (Vartiainen, Puska, et al., 1994) and mortality (Puska, Vartiainen et al., 1998), as well as reductions of cancer risk factors (Luostarinen, Hakulinen, & Pukkala, 1995).

Hospital-Based Change in Patient Perceptions and Behaviors Summative evaluation approaches that permit the researcher to observe the effects of

an intervention between or among groups randomized to different experimental conditions (McBurney & White, 2010) are common in clinic-based randomized controlled trials. The design involves multiple baseline measures that are often repeated over time. In such designs, each individual in a group is followed and measurements are collected over the period of the study in order to illuminate between-group changes over time. One of the primary advantages of this design is the increased statistical power afforded by removing subject variance; in short, within-person changes in responses in one condition can be directly compared to within-person changes in responses in another condition (Greenwald, 1976; Stangor, 2007). For example, a study to evaluate effectiveness of a brief intervention designed to alter patients' perceptions about their first MI utilized a prospective randomized design in which patients received the intervention or usual care from rehabilitation nurses. Patients were assessed in the hospital before and after the intervention and at 3 months after discharge from hospital. The intervention resulted in significant positive changes in patients' views of their MI, in feeling better prepared to leave the hospital, in reporting significantly lower rates of angina symptoms, and in returning to work faster than the control group (Petrie, Cameron, et al., 2002).

School-Based Adolescent Health School-Based Adolescent Health Care (SBHC) programs were intended to increase adolescents' access to a range of basic health services, to reduce the prevalence of high-risk behaviors, and to serve as demonstration projects to establish whether centers of that type could be established and run effectively in low-income urban communities. Participating organizations included public health departments, teaching and community hospitals, community health centers, and nonprofit community health agencies that operated SBHC programs in 24 junior and senior high schools in 19 communities with populations of 100,000 or more (Lear, Gleicher, et al., 1991). However, limitations curtailed the extent to which outcomes could be measured using random assignment of students to an experimental or control group. Evaluation options

were initially limited by the project timeline, in which the evaluation plan was only designed after the health centers began operation. In this way, the SBHC programs exemplified the tension that often exists between public health program delivery priorities and evaluation research agendas. As Knickman and Jellinek (1997) write, while a formal evaluation had always been planned, “the primary question for program planners was whether school-based clinics were viable on a broad scale: would diverse school districts take the risks necessary to get clinics up and running?” (p. 609). The clinics’ provision of sexual and reproductive health-care services had the potential to be controversial and, as such, the program staff focused mostly on designing an initiative that could be implemented in local communities. Only after the project was launched and the communities attempted to start their clinics were the program staff able to direct resources toward formal evaluation activities.

By that point a random assignment design was not possible since schools had already been selected (nonrandomly) without agreeing to the rigors of randomization. Random assignment of students within schools was also ruled out, both for the ethical concerns regarding withholding care and the practical concerns of spillover and potential contamination effects from intervention to the comparison group of students. Instead, the original evaluation design included a matched comparison sample of schools that had not opened health clinics in the same school districts as the SBHC sites. However, senior program leadership became concerned that student surveys about sexual behavior in the comparison schools could lead to parental backlash and undermine the support of the participating school districts to implement the clinics. This in turn could affect the outcomes of the primary evaluation question regarding the feasibility of the initiative. It was decided that surveys of student behavior would exclude comparison schools within the same district and would instead compare changes in high-risk behaviors among students in the schools with clinics to a national sample of urban youths (Knickman & Jellinek, 1997). To measure these outcomes, the research

design entailed two longitudinal surveys: one with the health center school students and the other with a national sample of urban youth in the same grades. These parallel surveys conducted over multiple time points gave the researchers a group with which to compare trends in behavior and outcomes. The primary limitation of this method is that such studies are less likely to detect smaller program effects, given natural variation across sites. However, given the challenges of the research, this approach offered credible (if not entirely conclusive) evidence on other program effects (Kisker & Brown, 1997; Knickman & Jellinek, 1997).

Can the Intervention Outcomes Be Replicated?

A replication study is a deliberate repetition of research procedures in a second investigation for the purpose of determining if earlier results can be confirmed and further supported (Polit & Beck, 2008). Replication of findings is one of the most powerful tools available to validate claims in scientific research. By helping to confirm or dispute findings of an original study, replication studies can also promote the generalizability of the original study or allow unsupported findings to be dropped from practice. In other words, investigators can conduct a replication study to see if the findings from an original initiative are applicable, or generalizable, to their population of interest. Although replication can be incorporated into primary study designs, such as those utilizing multiple baseline measures or replicating an intervention among wait-listed controls after a first wave of outcome evaluations are complete, replication studies typically take one of the following three forms.

Identical Replication

The first is an identical replication study in which the original study is repeated exactly with the same sampling procedures, measurement tools, and analyses. For example, in a replication of the Go Sun Smart program, a behavioral intervention focusing on sun safety behaviors of ski resort

employees and guests, the original research protocol was repeated at the sites that served as control groups in the original study. Using the same messages, measurements, and analyses, researchers were able to reproduce the results of the original study, in which greater exposure to intervention messages was associated with greater use of sunscreen, sunscreen lip balm, and face covering, but not gloves or overall sun protection (Andersen, Buller, et al., 2009).

Partial Replication

In the second type of replication study, an original study is duplicated as closely as possible, but not identically. Two major hand hygiene promotion interventions previously demonstrated to induce sustained improvement in clinical settings were replicated along with a passive intervention (soap substitutions and introduction of alcohol-based hand rub, with short-lived promotion of the changes) in selected wards of an 800-bed university teaching hospital. Each intervention used a before-and-after study design to assess results only within, not between, programs; the researchers chose this model because although all three interventions were conducted in parallel, there was no intention to compare them because potential confounders identified by previous modeling could not be controlled in statistical tests of significance. By replicating both successful interventions, the researchers were able to confirm that the programs can improve hand hygiene compliance and that the improvements can be maintained post-intervention. However, because the interventions were not delivered identically to the original study (e.g., implementation varied based on departmental engagement and leadership), the investigators were able to use differences in outcomes as compared to the original study to identify institutional support, commitment, and guidance as active ingredients in the success of the program (Whitby, McLaws, et al., 2008).

Systematic Extension Replication

The third kind of replication study is a systematic extension replication, which tests the implications of a study in a new setting to establish broad ecologic validity. For example, the initial

investigations into the links between procrastination and health in student samples implicated stress-related and behavioral pathways. Researchers who sought to replicate and extend previous findings among community-dwelling adults found that, consistent with previous work, procrastination was associated with higher stress, more acute health problems, and practicing fewer health-promoting behaviors (Sirois, 2007). Conversely, the positive results of an HIV/STI intervention originally targeting urban African-American males in nonschool settings were not successfully replicated in health classes at urban and suburban schools with diverse student bodies. The replication demonstrated increased knowledge, confidence, and behavioral intention among the intervention group but had no impact on sexual initiation, frequency of intercourse, or condom use, leading the investigators to conclude that the behavioral impact of an intervention may not be easily transferable when the program is taught to different groups and/or outside of the original setting (Borawski, Trapl, et al., 2009).

Challenges in Replication

Replication studies can pose challenges. First, one negative replication does not necessarily invalidate an original positive finding. Interpretation of the results of replication studies must take into account the myriad reasons that attempts to repeat the results might not be successful. Interpretations of failure must avoid the error of affirming the null hypothesis. Even if the failure of replication raises questions about generalizability, it cannot falsify the original finding. For example, if the original effect is small, negative results may arise by chance alone. Additionally, the participants or environment in a replication attempt might differ from those in the original study in a way that becomes significant, or a team might lack the skill or resources to reproduce the study correctly (Yong, 2012). Beyond these, numerous features such as those of settings, organizations, implementation fidelity, and populations addressed may moderate the generality of a finding from its original setting to others. Prudence and parsimony suggest always

assuming failure of generalization before concluding lack of validity of original findings.

A second source of challenge of replication studies is that, despite the value they may play in understanding behavioral interventions, it can be difficult to fund and publish replications, primarily because they are viewed as adding few novel findings to the existing literature relative to the time and resources spent on the research (Carpenter, 2012; Jasny, Chin, et al., 2011). In spite of virtually all authoritative writing on the subject identifying replication as essential to scientific progress, support for replication is very rare. This is in fact one of the challenges faced in much dissemination research, which is the systematic study of replications, the subject, in part, of our next section.

How Can the Intervention Be Improved, Scaled, and Disseminated?

A variety of methods and tools have emerged that offer practical assistance to help individuals and organizations make improvements in program delivery and outcomes and set the stage for dissemination and scaling (Duke University Health System, 2018; Langley, Nolan, et al., 1996; Moen, Nolan, & Provost, 1999). One of the most practical approaches utilizes *rapid cycle improvement*, in which the emphasis is on implementing small tests of change. These changes are based on ideas which might come from the process and outcomes evaluations, from the literature, from practices seen in other programs, or from new opportunities that emerge as structures evolve (either specific to the program level, or more broadly, such as the development of new technology and media).

Rapid cycle improvement initiatives often utilize the *Plan-Do-Study-Act* (PDSA) cycle. This begins with the “Plan,” which includes assembling a knowledgeable, motivated team to develop an *Aim Statement*. The Aim Statement articulates the specific, measurable goals to guide the improvement effort, as well as stated measurement objectives to determine whether the changes were effective. Aim Statements can

often be effectively developed by thinking broadly about change concepts, i.e., generic ideas that can be applied to spark a specific change in the situation. Change concepts might include managing time (e.g., reducing startup, setup, or wait time), avoiding mistakes, improving workflow, or minimizing waste.

Thus in PDSA, the plan emerges from trialing, making numerous small changes, and revising, rather than from a protracted planning exercise designed to arrive at some “perfect” plan. The “Do” phase involves carrying out the plan. The change may be tested with only a small number of patients/staff/program participants, and the test period may be as short as 1 day for small PDSA cycles. The *Study* phase involves examining the results to determine if objectives were met. All PDSA activities should be documented in detail to allow for comparison between different plans. Finally, *Act* uses the results to make decisions, incorporate changes into the workflow, and establishes future quality improvement plans. If the improvements were successful on a small scale, they should be tested on a wider scale to ensure an acceptable level of improvement is achieved. At that point, plans should be made to standardize the improvements. If the change was not an improvement, the team should develop a new theory and test it; often, several cycles are needed to produce the desired improvement (HealthIT.gov).

For example, a Michigan public health department utilized PDSA in an effort to improve its older adult influenza vaccine programs. The Aim Statement was, “Increase older adult (65+) influenza immunizations to achieve an 80% influenza immunization rate by the end of the next flu season” with three improvement outcomes measurements: (a) percentage of eligible persons who receive a vaccine; (b) increase in the percent of ordered vaccine that is administered; and (c) increase in the number of sites offering influenza vaccines to older adults. To understand the current vaccination processes, the PDSA team developed a flow chart of the major steps for the health department and community partners involved in annual influenza vaccinations, including ordering vaccine, scheduling clinics, and distributing and retrospectively evaluating the success of efforts.

This revealed that almost all steps were carried out independently, without coordination among the various entities who delivered vaccinations. Further, some community-wide program elements, like public information, were performed by the health department with no input from other providers. Finally, no one was aware of any process to evaluate what strategies were effective across the community while the flu season was in progress, and there was no mechanism within which to provide data or feedback to providers about their patient and community-wide vaccination rates. The team brainstormed potential solutions to these shortcomings, considering potential costs, potential impact, and feasibility of success of each idea. The most promising solutions were tested at 20 randomly selected providers, data were collected and evaluated, the changes to practice were deemed to increase vaccination rates, and the program was standardized and implemented with the full roster of providers the following year (Tews, Heany, et al., 2012).

Engaging the Community in the Research Process

Having reviewed a broad range of research and program evaluation methods to answer the seven questions from needs assessment to dissemination, we turn to a cross-cutting theme—the engagement of communities in program development, research, and evaluation. Researchers and practitioners are increasingly realizing that improvements in population health and the problems of interest in behavioral medicine require changes in a broad range of social determinants of health. Achieving changes in these challenging areas needs to include working with communities through partnerships between researchers, practitioners, and members of the community. Community-based participatory research (CBPR) has emerged in recent decades as a collaborative research approach designed to bridge the gap between science and practice through community engagement and social action to increase health equity by ensuring and establishing structures for participation by communities affected by the issue being studied (Israel, Schulz, et al.,

1998, 2001; Minkler, Blackwell, et al., 2003; Wallerstein & Duran, 2010). Some proponents of CBPR have come to use the term more under an umbrella of community organization approaches to intervention. Here we intend a broader use to describe the close relationship between researchers and communities that seek engaged understanding of the challenges those communities face and collaboration in developing responses to them—ranging from better individual clinical care and health education to broad public policy. (See also accompanying chapter by Ramanadhan & Viswanath.)

CBPR involves a reciprocal transfer of expertise, shared decision-making, power, and mutual ownership over the processes and products of the research (Freudenberg & Tsui, 2014; Viswanathan, Ammerman, et al., 2004). It expands the potential to develop, implement, and disseminate effective interventions across diverse communities through strategies to address power imbalances and facilitate mutual benefit among academic and community partners. Thus, at its best, CBPR can not only lay the foundation for efforts that improve population health, but also create broader community capacities for addressing issues that support improvements in other spheres of community development, including the environment, housing, transportation and economic activity, and in policy changes that create a just and humane society (Freudenberg, 1982; Freudenberg, Franzosa, et al., 2015; Freudenberg & Tsui, 2014). It also allows for practicing health professionals to engage in both an analysis and implementation of solutions unique to the specific setting, in collaboration with those who live and practice in that setting (Livingood, Allegrante, et al., 2011). Perhaps most importantly, CBPR encourages and promotes reciprocal transfers of knowledge (Wallerstein & Duran, 2010) by training community members in research (Minkler, Lee, et al., 2010) and including them in intervention development, e.g., the use of former drug users as “translators” in the design of a program to reduce sexual risk among African-American cocaine users (Stewart, Wright, et al., 2012), and delivery, such as through the use of local community volunteers to provide instruction on a physical activity program in Iran (Pazoki, Nabipour, et al., 2007).

The literature identifies several principles of practice to help community–research collaborations in developing, implementing, and evaluating their partnerships. These include the following: (a) identifying the best processes based on the nature of the issue and the intended outcome; (b) academic and community partners learning from each other; (c) capacity building (e.g., the commitment to training community members in research) (Minkler, Lee, et al., 2010); (d) acknowledging the difference between community input and active community involvement, and emphasizing the latter; (e) developing relationships based on mutual trust and respect; (f) acknowledging and honoring different partners’ “agendas”; (g) collaborating not only in applying findings but also in determining the ways in which the findings are produced and thus interpreted (Green & Mercer, 2001); (h) using evaluation strategies that are consistent with the overall approach taken in the academic–practice–community partnership; and (i) engaging in long-term commitments to effectively reduce disparities (Baker, Homan, et al., 1999; Green & Mercer, 2001; Wallerstein & Duran, 2006).

Several examples of CBPR from around the world demonstrate how these principles have guided various community health improvement efforts. The first example is of the Mayisha project, which involved a participatory community-based survey among five migrant sub-Saharan African communities in London. The research plan was guided by a community-based collaborative group selected to encompass broad experience in sexual health and HIV research and in HIV prevention with African communities, including representatives from African HIV forums, Client Care Services, Directors of Public Health, and African Health Promotion teams. Community fieldworkers from the local African community were identified through key stakeholders and local advertising and were responsible for recruiting participants to the study. From a practical standpoint, the use of community-based fieldworkers allowed the investigators to ethnically match interviewers to participants. From a larger public health perspective, the use and acceptability of participatory methods allowed the African communities to

demonstrate their commitment toward supporting studies of this nature and improving sexual health (Fenton, Chinouya, et al., 2002).

A second example is that of a project undertaken in Beirut, Lebanon. A 3-year CBPR project involving the testing of a psychosocial intervention to improve the reproductive and mental health of married women in a disadvantaged community of Beirut partnered university researchers with a community advisory committee and a local women’s committee. Evaluation of this approach found that the women and the broader community felt ownership of the study and that the CBPR approach gave the women voices and allowed for an improved understanding of the community and surrounding reproductive and mental health issues (Kobeissi, Nakkash, et al., 2011).

A third illustrative of CBPR collaboration in the United States began with a participatory, door-to-door health survey of 1000 households in New Castle, a small municipality located in rural Indiana, that revealed a smoking rate of twice the national average. This finding helped galvanize the community into action to develop and implement a variety of health-promoting environmental and policy changes, ranging from restrictions on indoor smoking to initiatives to promote physical fitness and healthier lifestyles. In this way, the CBPR approach laid the groundwork for long-term sustainable changes in support of community-wide improvements in health (Minkler, Vasquez, et al., 2006).

Finally, a CBPR approach that has been used successfully to address substance use prevention through school community health promotion in Iceland provides a good example of a long-term effort (Sigfusdottir, Thorlindsson, et al., 2009; Sigfusdottir, Kristjansson, et al., 2011). School-based surveys in the early and late 90s showed that substance use among 13–15-year-old adolescents was on the rise in Iceland. This led to development of a CBPR approach where academic and practice-based researchers, policy makers, and field practitioners in adolescent health, municipal leisure services, and education came together and collectively organized a program that continues to the present day. This program, referred to as the *Icelandic Model*, emphasizes

the school district as a unit of intervention, analysis, and reporting. Population surveys have been carried out annually, with input from all concerned stakeholders, including school personnel and community-based parent groups, and findings on mutually identified risk and protective factors have been analyzed and disseminated for each school district and municipal community taking part in the program. This approach has contributed to a paradigm shift in norms, values, and perceptions about adolescent health and development throughout Iceland. Moreover, during the 15 years that the approach has been ongoing, substance use among ninth and tenth grade students has dropped over 60% in Iceland (Kristjansson, James, et al., 2010; Sigfusdottir, Kristjansson, et al., 2008). In this case, CBPR has provided an empowering opportunity for communities to define and take ownership of a critical problem and find the practical and feasible solutions to address it.

In addition to benefits for the community, this CBPR effort over more than a decade has produced rich new findings illuminating key issues in the field, as well as advancing substance use prevention. These have included new methodologic innovations in conducting ongoing survey research (Kristjansson, Sigfusdottir, Sigfusson & Allegrante, 2014; Kristjansson, Sigfusson, et al., 2013) and investigating and addressing emerging community health challenges, such as the problem of physical inactivity (Eidsdottir, Kristjansson, et al., 2008) and overweight and obesity (Eidsdottir, Kristjansson, et al., 2010, 2013; Thorisdottir, Kristjansson, et al., 2012). The collaborative research has also increased understanding of the relationship of body mass index and depressive symptoms in adolescents (Eidsdottir, Kristjansson, et al., 2014).

For all of its potential to strengthen the collaboration between researchers, practitioners, and lay members of communities to solve health problems, CBPR also poses some challenges. First, logistically, behavioral medicine investigators who adopt the principles of CBPR in their work with communities can face obstacles in several areas: partnership capacity and readiness, time constraints, funding flexibility, translation, and expansion (Macaulay et al., 2011;

Minkler, Blackwell, et al., 2003). Second, CBPR can also require researchers to confront and address a range of thorny ethical issues, including how to obtain participation and community consent, tensions created by differentials in power and privilege, racism and ethnic discrimination, and a range of issues around research for social change (Green, 2004; Wallerstein & Duran, 2006). Finally, respect for the community needs to be combined with an understanding that *all human beings* are often inaccurate in their explanations of their own behavior. For example, individuals entering smoking cessation classes often request dramatic portrayals of the harms of smoking (e.g., photographs of blackened lungs) as an aid in their efforts to quit. The role of such “scare tactics,” however, is complicated. They may encourage progress in the early stages of contemplating quitting (Hammond, Fong, et al., 2004) or predict quit attempts (Yong, Borland, et al., 2014), but evidence for their impact on actual cessation is still lacking (Borland, Yong, et al., 2009). In one study of different types of support for adults seeking to lose weight, that which received the highest satisfaction ratings was not that in which participants achieved the greatest weight loss (Gabriele, Carpenter, et al., 2011). It seems that collaboration and mutual respect in program planning need to include recognition of these features of the psychology of self-perception, social influence, and attributional processes along with respect for the perceptions of those involved and evidence regarding effective approaches. Despite these potential pitfalls, familiarity with the concepts and principles and the spirit of CBPR, together with good process evaluation, can help investigators and program planners avoid them.

Translating Research into Practice

An emerging area of concern related to the issues of replication and dissemination is the translation of research findings into actionable practice and what has been sometime referred to as the “gap” between research and practice. For example, in the case of community-level efforts

to prevent injury, Hanson, Finch, Allegrante, and Sleet (2012) have identified three principal gaps that separate academic researchers, policy makers, health practitioners, and the communities in which change is being proposed. These include: (a) the research-to-practice gap, (b) the efficacy-to-effectiveness gap, and (c) the injury-prevention-to-safety-promotion gap (Hanson, Finch, et al., 2012). In reviewing over 1200 articles published in 12 leading public health and health promotion journals, Oldenburg, Sallis, Ffrench, and Owen (1999) found that 63% of publications were descriptive, 11% were concerned with method development, and 16% were intervention based; only 5% were concerned with institutionalization or policy implementation research, and less than 1% contained diffusion research. This is important for behavioral medicine because it is not at all uncommon for interventions that have been tested under conditions of high internal validity to be altered when they are implemented into practice settings (Cohen, Crabtree, et al., 2008). The literature has sought to address the challenges of translating research into practice (e.g., see Cohen, Crabtree, et al., 2008; Katz, Murimi, et al., 2011) and several approaches to improving the translation of research into practice have been proposed (e.g., see Glasgow & Emmons, 2007). In the following, we review several of these models.

RE-AIM

One of the most widely respected approaches to improving research translation in behavioral medicine research is the *RE-AIM framework*. The RE-AIM framework is one of the primary tools designed to enhance the quality, speed, and public health impact of efforts to translate research into practice. The goal of RE-AIM is to draw attention to essential program elements, including external validity, which can improve the sustainable adoption and implementation of effective, generalizable, evidence-based interventions (Glasgow, Vogt, & Boles, 1999). The five RE-AIM steps for evaluating the potential health impact of interventions are:

1. **Reach:** The absolute number, proportion, and representativeness of individuals who are willing to participate in a given initiative, intervention, or program.
2. **Efficacy or Effectiveness:** The impact of an intervention on important outcomes, including potential negative effects, quality of life, and economic outcomes.
3. **Adoption:** The absolute number, proportion, and representativeness of settings and staff or other people who deliver the intervention (i.e., intervention agents) who are willing to initiate a program.
4. **Implementation:** At the setting level, implementation refers to the intervention's fidelity to the various elements of its protocol, including consistency of delivery as intended and the time and cost of the intervention. At the individual level, implementation refers to clients' use of the intervention strategies, consistency, costs, and adaptations made during delivery:
5. **Maintenance:** At the setting level, maintenance refers to the extent to which a program or policy becomes institutionalized or part of the routine organizational practices and policies. Maintenance also applies to intervention effects at the individual level over time, defined as effects of a program on outcomes after 6 or more months after the most recent intervention contact.

RE-AIM's utility and effectiveness derive in large part from its focus on both individual levels (reach and efficacy) and organizational levels (adoption and implementation) of impact (e.g., maintenance can be both an individual- and an organizational-level impact). It is critical to evaluate both levels because each provides valuable independent information about intervention impact. For example, a clinic-based intervention that has large impact on reach and efficacy but is only adopted, implemented, and maintained at a small number of medical practices with specific resources that are not available in typical "real-world" settings would not have sustainable translation into wider practice. If only the individual dimensions of the intervention were used for evaluation, the intervention would appear to have

large potential for impact when in reality it has little hope of resulting in a large public health impact because it could not be adopted, implemented, and maintained in real-world settings. Conversely, if an intervention has potential for wide organizational adoption, implementation, and maintenance, but little reach, efficacy, or maintenance at the individual level, the potential impact of the intervention would not likely be achieved because of the deficiencies at the individual levels.

The RE-AIM website (www.re-aim.org) maintains a substantial searchable library of publications utilizing the framework. Examples include the planned evaluation of the BETTER 2 program, which is designed to expand implementation of a chronic disease prevention and screening intervention in primary care settings. Researchers will evaluate the program using RE-AIM to inform a mixed-methods approach, including a descriptive statistics on patients accepting the intervention, qualitative information on implementation and adaptations of the program, longitudinal measures of provider use of the intervention within their practice, and a composite index to assess quantitatively the effectiveness of the intervention (Manca, Aubrey-Bassler, et al., 2014). Likewise, an evaluation of different versions of an Internet-based diabetes self-management support program provides a clear application and interpretation of the RE-AIM model (Glasgow, Kurz, et al., 2010). A three-arm practical randomized trial was used to compare a minimal contact and moderate contact versions of the online program, compared to an enhanced usual care protocol. Primary behavioral outcomes (e.g., healthy eating, physical activity) and secondary biological outcomes (e.g., hemoglobin A1c, BMI) were compared at baseline and a 4-month follow-up. Interpreting the results through the RE-AIM structure, the intervention met several of the criteria for potential public health impact, including that it was feasible and engaging for participants and was able to reach a large number of people. However, there was mixed effectiveness in improving outcomes, and the authors concluded that further research is necessary to evaluate long-term outcomes, to

enhance the effectiveness and cost-effectiveness of the intervention, and to better understand the connections between the intervention's processes and its outcomes. In this way, while RE-AIM is most commonly used to report results or compare interventions, it is also useful as a planning tool as well as a method to review intervention studies.

Utilization-Focused Evaluation

Another approach to facilitating translation of research into practice is what has been referred to as *utilization-focused evaluation*. Evaluation researchers must bear in mind that intended users of research, such as community leaders, program managers, policy makers, and public health-related practitioners, are more likely to use findings from evaluations if they understand the research process, are consulted and engaged, and sense ownership of the findings. This approach is the basis for utilization-focused evaluation, a process for making evaluation decisions in collaboration with an identified group of users, with a focus on their intended uses of the evaluation. By actively involving users in this way, the evaluation researcher can prepare the ground work for the use of evaluation findings and train users in the use of findings, which reinforces the intended utility of the evaluation throughout the process (Alkin, 2004). The approach thus places priority on how the evaluation findings will be applied by people in the real world (Patton, 2008).

Utilization-focused evaluation does not advocate any particular evaluation content, model, method, theory, or use. Instead, it is a process for identifying key stakeholders (Bryson, Patton, & Bowman, 2011; Guba & Lincoln, 1989) and helping these primary intended users to select the most appropriate content, model, methods, theory, and uses for their particular situation. This process has been employed to evaluate a wide range of topic areas, including human services agencies (Greene, 1987), a national AIDS prevention program in Switzerland (Dubois-Arber, Jeannin, & Spencer, 1999), and a Canadian compassionate

care benefit (Williams, 2010). Along with several guides and toolkits on the topic (Bryson, Patton, & Bowman, 2011; Fetterman, 2000; Patton, 2008, 2012), investigators interested in the logic of the utilization-focused evaluation approach might find it useful to compare the pretest-posttest control group design used to evaluate a problem-solving skills training for adolescents (Tellado, 1984) to the alternative, utilization-focused research plan proposed for the same program (Patton, 1984).

PRISM

A third promising approach that supports translation of research into practice is the Prevention Impacts Simulation Model (PRISM). Originally designed and developed at the US Centers for Disease Control and Prevention and National Institutes of Health to estimate the impacts of public health interventions on the health of populations, PRISM is a systems dynamic simulation tool that includes 22 categories of policy, systems, and environmental change across several broad areas of interventions (Homer, Milstein, et al., 2008, 2010; Honeycutt, Wile, et al., 2014). These categories address medical care, smoking, nutrition and weight loss, physical activity, emotional distress, and air pollution. PRISM can assist users with decision-making about chronic disease intervention by modeling the likely impact of various prevention strategies for cardiovascular diseases, as well as for cancer and respiratory diseases related to risk behaviors such as smoking, diet, and physical inactivity. Moreover, PRISM is capable of modeling both individual interventions and combined interventions and can be used not only to estimate the short- and long-term population effects of intervention and the future costs averted by prevention but also to forecast future outcomes. By integrating the best available evidence on modifiable risk factors and demonstrated interventions, PRISM thus can provide a robust simulation of the impact of proposed or implemented prevention efforts on both indicators and cost of chronic disease.

Emerging Issues and Challenges

The movement toward evidence-based behavioral medicine practice mirrors one that has evolved in clinical medicine over the past several decades. Encouraged by health-care reform initiatives around the world, this movement has been fueled by a rapidly growing foundation of evidence from outcomes research in clinical medicine, behavioral medicine, and public health. Perhaps the most visible demonstration of the evidence-based medicine movement is embodied in the Cochrane Collaboration (2015), which was formed to organize the burgeoning volume of medical research into a searchable and useable evidence base that could benefit decision-making by health professionals, patients, and policy makers. However, with this growth has come a spirited discussion among practicing interventionists, behavioral scientists, clinicians, evaluators, policy makers, and patients about what kinds of evidence actually constitute “the best available scientific data” to inform decision-making in clinical practice. What evidence is most likely to advance behavioral medicine, clinical care, and public health? What are the ideal channels and practices for the dissemination of new knowledge, and for translating research to best practices in real-world settings? And what ethical obligations does the scientific community have to ensure the widest possible dissemination of the benefits of research? While these questions are neither exhaustive nor limited to the research methods and approaches addressed in this chapter, they are nevertheless illustrative of the kinds of challenges behavioral medicine faces going forward. Thus, this final section of the chapter takes up several of what we believe to be some of the most pressing issues these and other questions raise for the field.

What Constitutes Evidence?

Even if in behavioral medicine we limit our investigations solely to pursuing improvements in health, the objectives of behavioral research range widely from proving that some specific therapeutic event or intervention results in some specific

outcome, to showing that a set of community-based activities will reward government support of those activities as a means of preventing disease or promoting health. Thus, the idea that any one approach to research would constitute “the gold standard” by which to answer questions across such a wide and variegated range of potential research objectives is preposterous. One need only ask what Copernicus, Darwin, Galileo, and Watson and Crick—all great scientists—have in common? *None of these great scientists had control groups!*

With growing recognition of the impact of health policies not only on health status, mortality, and quality of life but also on national economic prosperity and security, there is growing interest in using research to inform decisions with broad policy implications on sound evidence. However, concern over the soundness of evaluations and the stakes involved has encouraged some conservatism in ascribing value to the broadest range of available data. For example, a number of reimbursement policies in the United States now use evidence reviews of the [U.S. Preventive Services Task Force](http://www.uspreventiveservicestaskforce.org) (www.uspreventiveservicestaskforce.org) as the basis for identifying reimbursable services; however, the task force tends to confer considerable weight on evidence generated from RCTs and discounts the evidence from other equally valuable (but not equally valued) research designs. To illustrate the dilemma, consider a report in *The New York Times* (Kolata, 2013) on recent American Heart Association cholesterol guidelines. The story pointed to how “the drafting committee mistakenly relied only on randomized controlled clinical trials, the gold standard of medical evidence, but ignored other strong data [including a wealth of existing genetic and population data] that would have led to different conclusions.” This example demonstrates the risk of limiting the evaluation of interventions best evaluated through methods other than RCTs, and highlights the need to identify practical ways of identifying promising approaches to prevention and improved care in whole populations.

Thus, it is not surprising that a primary criticism of the RCT focuses on the need for realis-

tic, pragmatic alternatives to the dominant paradigm of studies requiring “hard” data and statistical proof utilizing highly homogenous patients in academic clinical settings. Such studies may be high in internal validity but they lack the external validity to support effective translation of research findings into policy and real-world practice (Glasgow, 2008, 2013; Steckler & McLeroy, 2008). Differences in culture, social structure, norms, and functions of communities and their populations, moreover, naturally preclude the highly widespread generalizability of results between populations affected by different social determinants (Livingood, Allegrante, et al., 2011). In addition, application of interventions previously demonstrated to be effective using strict research protocols (where internal validity is highly valued) often fail to produce the desired effect in real-world settings (where external validity is highly valued). In some cases, the intervention is initially delivered by trained professional study staff that is highly supportive of patients and motivated toward the success of the study. However, unless the intervention is one that can be easily replicated or continued in the hands of other staff members, interventions often fail to be maintained once study staff leave (Glasgow, Bull, et al., 2002). Moreover, even when programs are successfully continued by site staff beyond the study period, and the intervention is applied in a rigorous manner with high fidelity to the original study design, practitioners may find that the previously successful program is ineffective and fails to produce significant treatment effects (Hallfors, Cho, et al., 2006).

This problem is now sparking development and rapidly growing respect for innovative research and evaluation designs that allow interventions to gain credence on the weight of converging evidence and without the inherent limitations of the RCT. A good example of this movement toward more innovative thinking in evaluative research is the recent Institute of Medicine (IOM, 2012) study of how best to assess the value of community-based approaches to health promotion that go beyond the RCT

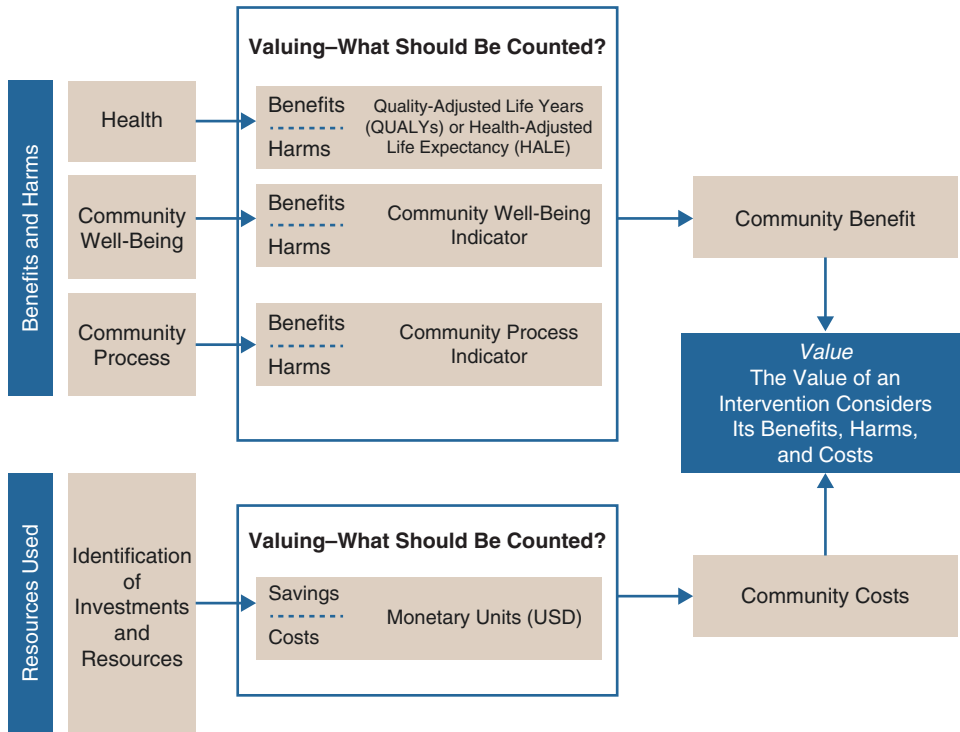


Fig. 6.1 Conceptual framework for assessing the value of community-based prevention interventions (Reprinted with permission from the National Academy of Sciences, Courtesy of the National Academies Press, Washington,

DC. A full copy of the report, *An Integrated Framework for Assessing the Value of Community-Based Prevention*, is available at: http://www.nap.edu/catalog.php?record_id=13487)

model of evaluation. The IOM report provides a new conceptual framework (see Fig. 6.1) for assessing the value (which includes the benefits, harms, and costs) of community-based prevention. The report points to the clear need to use constructs beyond individual health in assessing the value of health promotion and public health, including community well-being and community processes, and how public health and health promotion can more effectively assess and incorporate the dimensions of social determinants into programs and interventions. This will require the development of new performance measures, new metrics that will be capable of operationalizing the concepts of community well-being and community process, and ultimately new concepts of community benefit, as well as systematic ways of using the new framework as recommended in the IOM report (Allegrante & Livingood, 2013).

What Are the Objects of Intervention That Explain Variance in the Observed Effect?

A second question of interest has to do with what the objects of intervention (also sometimes referred to as the “active ingredients”) that are responsible for and explain the variance in the observed effect. To illustrate, evaluating the effect of a pill that contains a powerful medicine is very different from evaluating a national campaign to reduce obesity and type 2 diabetes. Especially important is that in assessing the effects of the pill, we seek to isolate the effect of the pill from the “confounding” ecologic effects of the context. This is why we conduct RCTs, which allow the investigator to isolate the active ingredient of intervention under highly controlled conditions. For a national campaign to reduce obesity and type 2 diabetes, however, we seek to

understand, incorporate, and exploit the effects of context, which interact with intervention components (often synergistically) and will change over time and space, as an integral part of the intervention. Thus, a primary issue for behavioral medicine research concerns the identification of key components of interventions to which outcomes can be attributed and, then, how to disseminate such interventions.

The experience and results of the Diabetes Prevention Program (DPP) illustrate the challenge. The DPP showed that moderate physical activity (150 min per week) and loss of 7% of body weight among those with impaired glucose tolerance significantly reduced conversion to type 2 diabetes in this high-risk group, both in comparison to placebo and to Metformin, a standard medication used to treat diabetes. This “lifestyle” intervention utilized a standard set of materials and a combination of group and individual contacts to promote physical activity and weight loss. However, in the interest of testing the benefits of physical activity and weight loss, *not* any particular way of achieving them, study sites were encouraged to be highly flexible and creative in developing ways that would enable participants to achieve 150 min of moderate activity per week and 7% weight loss. The exciting findings of the DPP, in many respects the greatest impacts identified to date for a fundamentally behavioral approach to a major health problem, raise the question of what precisely should be disseminated. The benefits of 150 min of moderate activity a week and 7% weight loss should lead to diverse efforts to achieve these; however, some voices have called for focus on the particular intervention used in the DPP, “a proven, community-based weight loss program,” i.e., “a...program,” not multiple approaches to pursuing weight loss and physical activity as the path to saving Medicare \$7 billion (Thorpe & Yang, 2011). There is a difference in strategy here: (a) disseminating a “program” that is as well and specifically defined as possible or (b) promoting varied approaches to achieving behavioral impacts, weight loss, and increased physical activity. It is likely that both strategies will have their successes. The difference between them

illustrates the challenges interventionists face in recognizing the need for both standardization and flexibility or adaptability in the implementation of programmatic elements of intervention.

What Are the Factors We Need to Study Not Only to Guide Dissemination but Also Adaptation and Implementation? And How Do We Incorporate Them into Research?

Ideally, no investigation should be limited to the researchers completing their project, publishing their manuscript, and moving on to the next project, leaving lessons learned buried in the archives of journals, or worse, not published at all. Finding ways to disseminate effective, affordable, and feasibly scalable interventions that have been demonstrated in research is a major challenge for behavioral medicine as it is for all population health professions. Dissemination should be guided in part toward objectives of sustainability and replicability by communicating to the scientific community not only the results and end point of the project but the explanatory details that drove the findings. This requires proactive consideration of dissemination goals before the program begins as well as meticulous documentation of all steps in the project before it concludes.

In what have come to be called *practical trials*, external validity and replicability are increased through four key features: (a) use of representative patients to represent the range of patients encountered in real-world settings; (b) implementation in multiple settings rather than in those that are expertise- and resource-rich; (c) comparison of conditions that represent current standards of care or alternative treatments (rather than placebo or no treatment) to demonstrate that changes in practice result in significantly better results than current, familiar interventions; and (d) the inclusion of multiple outcomes, such as implementation requirements, costs, and feasibility (Glasgow, 2008). Such trials can provide a much more complete picture of evidence-based practice and strengthen program delivery, but will only do so if the individual study methods and the intervention

are reported with sufficient detail and clarity to be fully understood and fully replicated.

In recent years, there has been recognition that translation of evidence into practice will be improved when research design and reporting standards are modified to help quality improvement teams understand both these adaptations and the effort required to implement interventions in practice (Cohen, Crabtree, et al., 2008). Toward this end, and as research knowledge has been accumulated, increased emphasis has been placed on the standardization and comparability of study findings that pertain to practice. A recent example of this is the Transparent Reporting of Evaluations with Nonrandomized Designs (TREND) Statement that provides a proposed 22-item checklist for standardized reporting of behavioral and public health intervention evaluations (Des Jarlais, Lyles, et al., 2004). The goal of this initiative is to increase the utility of nonrandomized study designs by providing a framework within which comparable information across studies can be more easily synthesized and translated into generalizable knowledge (Des Jarlais, Lyles, et al., 2004). It is intended to be a complement to the 25-item CONSORT Statement for randomized trials (Altman, Schulz, et al., 2001; Begg, Cho, et al., 1996) and the CONSORT-SPI (Montgomery, Grant, et al., 2013) that is designed specifically to enhance reporting clarity for social and psychological interventions. While TREND was originally launched to improve public health practice, behavioral medicine also stands to benefit from such transparency in reporting design and methods of quasi-experimental behavioral research. In addition to CONSORT and TREND, similar standards have now been developed for reporting qualitative studies (O'Brien, Harris, et al., 2014) and systematic reviews (IOM, 2011) and meta-analyses (Moher, Liberati, et al., 2009, for the PRISMA Group), along with other ongoing and emerging efforts that have been developed in the last decade to increase the quality and transparency of reporting in health research (Altman, Simera, et al., 2008) and reduce bias in randomized controlled trials (Higgins, Altman, et al., 2011).

Despite such guidelines and “rules,” no method of improving rules for gathering and transparency in reporting evidence can eliminate the role of scholarly, professional, or clinical judgment in assessing and applying it. Put simply, there are two broad tasks in this: gathering and assembling evidence and applying it. Perfecting the former does not eliminate the latter. That is, perfecting the evidence does not eliminate the role of human appraisal in judging how and to what it is best applied. Even when we can ensure that policy and practice will be guided by sound evidence, the role of human judgment in applying it remains fundamental.

How Do We Incorporate a System-Wide or Population-Based Approach to Research?

Since the WHO Report on the Social Determinants of Health (Commission on the Social Determinants of Health, 2008) demonstrated the powerful impact of economic, environmental, and social- structural factors on health status, more research effort has focused on these broader determinants. However, much published research has continued to focus on individual determinants of change, largely ignoring the other contexts that shape behaviors (Glass & McAtee, 2006; Golden & Earp, 2012). Broader trials are needed that include impact and outcomes important to decision makers and communities, that address multiple contexts outside the individual level, including the environment in which programs are conducted, and that focus on moderating and mediating factors, economic issues, and social contexts (Glasgow, 2008; IOM, 2012).

Several have called for behavioral medicine and public health to embrace behavioral interventions that are system-wide, population-based, or focused on changes in public policy, which by their very nature require consideration of a broader range of acceptable “evidence” and “outcomes” as well as increased inclusion of the varied contributors to public health (Allegrante & Livingood, 2013; Epping-Jordan, 2004; IOM, 2012; Lieberman, Golden, & Earp, 2013; Livingood, Allegrante, et al., 2011). Limiting the field to individually

based, “best practice” interventions for which there is “scientific evidence” from randomized trials not only fails to recognize the key social, policy, media, and other ingredients that actually produce significant results but can also point practitioners in the misguided direction of higher-cost, clinically based interventions (Livingood, Allegrante, et al., 2011). For example, Livingood, Allegrante, et al. (2011) note that, historically, major public health achievements tend to involve a complex and dynamic interaction between society and community rather than following “a linear movement from scientific testing to broad application” of individually focused interventions that have been characteristic of the classic biomedical model. The biomedical model, most notably developed by the US National Institutes of Health, comprises a model of scientific discovery on a continuum that goes from basic and clinical research, to applied research and development, to treatment at the bedside (Levy, 1982). In contrast, key public health achievements of the twentieth and twenty-first centuries, such as the normative changes in tobacco use in many countries, have required attention to a range of individual, social, and institutional factors that influence population health and have been the focus of health promotion (Livingood, Allegrante, & Green, 2016).

Reflecting this “range of individual, social, and institutional factors,” recent developments in statistical analysis techniques collectively referred to as *multilevel methods* now allow researchers to design studies that can disentangle individual-level influences from community-level factors such as class, school, work site, residential location, town, city, or county, but still examine all levels at the same time and thus come closer to “clean” effects that may be attributed to different levels (Raudenbush & Bryk, 2002; Snijders & Bosker, 2012). The essential idea with multilevel studies is that findings pertaining to individuals that are close in proximity for any number of reasons (e.g., live in the same neighborhood, attend the same school, and employed at the same workplace) are likely to be caused, at least in part, by similar lifestyles and/or living conditions. Multilevel inquiries thus uphold that research findings are not simply attributable to individuals but also to the social environment

and common living circumstances, which in turn may increase the strength of the investigator’s interpretation.

The Weight of Evidence and Affirming the Null Hypothesis

Research scientists are by nature skeptical and critical. As a consequence, they are frequently dismissive of propositions with the response that “*there’s no evidence for that*”. Technically, the refrain, “*there’s no evidence for that*,” allows the conclusion of just that, “there’s no evidence.” However, it frequently leads to an assertion that a particular line of inquiry or endeavor has been shown not to work—this amounts to “affirming the null hypothesis,” something most graduate students were taught was not possible. This can grossly constrain development of behavioral medicine and public health interventions of all kinds. The fear of “there’s no evidence for that,” should not constrain our creativity in developing new and untested approaches (Fisher, 2008).

There are additional problems in weighing the presence and absence of evidence. The law of parsimony, Occam’s razor, dictates that science assume the simplest relationship among events until evidence forces more complexity (Fisher, 2008). In light of this, consider, for example, findings of the Guide to Community Preventive Services that sufficient evidence exists for diabetes self-management education in “community gathering places” such as “community centers, libraries, private facilities (e.g., cardiovascular risk reduction centers), and faith institutions” (p. 201) but not if offered through worksites. (Only one study was found reporting self-management education in worksites and it “had design limitations,” p. 207). The variability in available evidence from different sites does not constitute evidence that there are significant or substantial differences among them. Parsimony would lead to a conclusion that there is evidence for diabetes self-management education settings in a variety of settings without any strong evidence to differentiate among them in terms of likelihood of success.

Recalling the inebriated individual looking for keys under the streetlight, not because that is where they were lost, but because “the light is better here,” we need to be critical of the assumptions we make in assembling evidence (Fisher, 2008). Why disaggregate diabetes self-management education according to community centers, libraries, private facilities, cardiovascular risk reduction centers, faith institutions, and worksites? To what extent do the distinctions among them have a plausible causal role justifying their differentiation? Why not large versus small settings? Daytime versus evening? Clinic- versus community-based? In contrast, perhaps more useful answers would flow from disaggregating by such factors as organizational support for the program, proximity or accessibility to intended audiences, and presence of community resources supporting the program (e.g., safe, enjoyable sites for physical activity). The point here is not just about community sites for diabetes education, but that how we assign evidence to groups or categories and, more broadly, how we manage and categorize evidence will influence the conclusions we draw from it. Parsimony dictates we do not differentiate without evidence. Greater criticality about these processes needs to come before “there’s no evidence for that.”

Another way to view the dilemmas is to imagine you were in charge of a state department or ministry of health and you were looking for evidence to guide public investment in population-wide prevention of diabetes. In weighing the available evidence to inform your decision, would you prefer to base your decision on 20 RCTs, all of which demonstrated exquisite internal validity showing that a particular approach was effective relative to randomized controls among volunteer samples treated through university research centers? Or would you prefer to base your decision on the findings from practice-based evidence of ten programs testing varied adaptations of an approach and associated community health promotion activities, carried out with urban, rural, and multiethnic community groups, and showing benefits in pre-post analyses and against national norms; associated lessons learned identifying local buy-in from a government or health leader, inclusion of

primary care endorsement, and duration of community activities as critical success factors? The point is that evidence can come in many forms and limiting our confidence to evidence generated solely from highly controlled conditions narrows the range of evidence from which we can draw in making decisions about what programs work and under what conditions. Ultimately, the example also points to Green’s (2008) notion of the “fallacy of the pipeline” that he argues is implicit in the traditional, unidirectional continuum of translation and dissemination of research into the hands of practitioners who are then expected to implement approaches that have been tested under condition of high internal validity. To counter this fallacy, Green has captured the challenge in what is a now-popular refrain: “If we want more evidence-based practice, we need more practice-based evidence” (Green, 2006, 2008).

Using PRECEDE-PROCEED, RE-AIM, PRISM, and similar models, if we identify needs related to a problem, identify reasonable evidence-based (where pertinent evidence is available) approaches to addressing those needs, implement the program and show it was implemented according to objectives, assess short-term impacts (e.g., reported changes in diet and physical activity), and show objectively measured reductions in the problem relative to appropriate benchmarks, does this not constitute knowledge with utility? This question is of distinct importance for the field of behavioral medicine in light of the urgency with which health disparities must be addressed. For example, in the United States most new HIV infections among youth occur among gay and bisexual males, with a 22% increase in estimated new infections in this group from 2008 to 2010 (Centers for Disease Control and Prevention, 2014). The timeline for an NIH grant to complete a large-scale randomized trial of a behavioral intervention can include 1–2 years devoted to securing funding and 3–5 years (or longer) devoted to conducting the research, completing data analysis and disseminating results, with the potential for the public health impact of the epidemic to increase significantly during that 7–10-year period. Further, this timeframe is particularly

inappropriate given the increasing role of new technology as tools for health promotion, for instance the HOPE (Harnessing Online Peer Education) social media intervention which increased HIV testing rates among young men via Facebook outreach and peer education (Young, Cumberland, et al., 2015). Accelerating the discovery, dissemination, and implementation of knowledge is thus a critical imperative.

What Are the Moral and Ethical Obligations of Dissemination?

Finally, no treatment of the topic of the application of research methods from education, applied psychology, and behavioral science to behavioral medicine would be complete without some consideration of the moral and ethical obligations of the researcher. What are our moral and ethical obligations as reflective scientists—whose work is often supported by public funds—to ensure that our scientific work and findings are disseminated? As noted previously, intended users are more likely to accept the utility of research and evaluation results, and more likely to support the dissemination of results, if they understand and feel ownership of the process of the research. This raises two significant issues.

First, beyond publication in peer-reviewed scientific journals, researchers increasingly need to dedicate adequate resources to community-based debriefings, discussion of research findings, and consideration of the implications for practical use. Researchers thus must be prepared to play the role of public intellectual to use their research in affecting change. All partners should be involved in the dissemination of information about the partnership and project findings in forms that all partners can understand and use; this includes reaching multiple audiences (e.g., community members, policy makers, local health professionals, and the lay public) through a variety of communication channels and formats (e.g., radio, newspapers, social media, presentations, handbooks, position papers, testimony, and scientific journal articles), with all partners involved as co-authors and co-presenters to the extent that substantive contribution to the research may require.

Thus, it is important to strike a balance between time devoted to the preparation of manuscripts for publication in peer-reviewed scientific journals and developing processes that enable researchers to report results to the broader community and public stakeholders for discussion of the utilities and implications of the research (Seifer, 2006).

Second, many researchers are now considering the ethical implications of publishing in peer-reviewed scientific journals where the published article will be housed behind a “paywall”—accessible only to readers whose academic institutions maintain costly annual subscriptions, or to those who are willing and able to pay a fee for access to a published piece of scientific research. Some critics charge that research funded by public tax funds should be freely available to the public whose moneys have supported the research, while others make the case that open access is a moral issue and that the principle of beneficence obligates scholars to act for the benefit of others. The open-access policies of the US National Institutes of Health (National Institutes of Health, 2014), the Canadian Institutes of Health Research (Canadian Institutes of Health Research, 2014), and the Research Councils, Innovate UK, and Research England of UK Research and Innovation (UK Research and Innovation, 2018) represent notable mandates adopted in the past decade which endorse this philosophy. Nevertheless, the issue of the extent to which the privatization of scientific content undermines the advancement of sciences remains a challenge, with important questions remaining about the system of scientific peer review and publication in which considerable knowledge remains out of reach for much of the general public.

Finally, the argument for open access has also extended to the issue of whether scientific peer review should also be open to broader audiences, including the public, rather than limiting judgments about scientific worth solely to submitting authors, journal editors, and anonymous reviewers. In 2006, the prestigious scientific journal, *Nature*, initiated a trial of *open peer review*. Open peer review is the process of rendering scholarly judgments about the scientific value of research through an entirely transparent process by which the identities of those reviewing the research are disclosed to submitting authors and the public as

part of the process. This differs from the traditional peer review process in which the identities of reviewers are anonymous. Despite significant interest in the trial by participating scientists, only a small proportion of those authoring papers chose to participate in the open review. The trial suggests that opening up peer review to broader participation among scientists and the public to comment on the quality and rigor of scientific research may not be as widely popular as believed (*Nature*, 2006).

Conclusion

This chapter has sought to describe and place in context the research approaches of education, applied psychology, and behavioral science and their potential applications to behavioral medicine. We have also attempted to sensitize the reader to the key issues in research and evaluation that will continue to require attention in behavioral medicine. Several final observations are worth making.

First, as we hope the chapter has shown, a wide range of research methods and evaluation designs are available to support the assessment of needs, formulation of intervention approaches, and evaluation of the process, impact, and outcomes of behavioral medicine interventions at the individual, community, and policy levels. Moreover, with the growing recognition of the importance of the broad range of social determinants of health, there is a nexus of complex factors that must be addressed to improve population health. To meet the challenge, more emphasis is now being placed on evaluative designs that value the use of a wide range of methods to assess health outcomes and the clinical, community, and social circumstances that support improvements in health and quality of life. Thus, the focus of this chapter has been on diverse methods in educational, public health, and behavioral science evaluation that may serve a broad range of objectives in order to study and understand complex phenomena and develop effective interventional measures.

Second, there are real and important differences in how interventions are evaluated that have profound implications for what and how

much we can learn from the research and programs in which society invests public resources. Simply because research may fail to document a prespecified or a priori outcomes of primary interest, does not necessarily mean a study is not without merit or that something valuable has not been learned. Even “negative” trials of interventions, or studies in which unanticipated outcomes of value have been observed, can yield vitally important insights and new knowledge that can guide others in future work.

Finally, behavioral medicine faces an evolving landscape of issues that will require its scientists and practitioners to work more effectively to engage the community and other stakeholders in designing and conducting investigational studies and program evaluations. Partnerships with other disciplines and with the patients and communities of interest whose health and quality of life we seek to improve promise to strengthen the science of behavioral medicine and the impact it can have through the research and evaluation process.

Acknowledgments We thank Edwin Fisher at the Gillings School of Global Public Health of the University of North Carolina at Chapel Hill and Linda Cameron at the University of California, Merced, for the helpful comments, attention to detail, and expert editorial guidance they provided throughout the writing of this chapter. We also thank Naomi Arbit, Madhabi Chatterji, Alexandra DeSorbo, and Miguel Marshall at Teachers College, Columbia University; Aduragbemi Banke-Thomas at the Liverpool School of Tropical Medicine; and Allyn Aringazina, MD, at the Kazakstan School of Public Health for reading and commenting on drafts.

References

- Abel-Smith, B., & Rawal, P. (1992). Can the poor afford ‘free’ health services? A case study of Tanzania. *Health Policy and Planning*, 7(4), 329–341.
- Aday, L. A., & Cornelius, L. J. (2009). *Designing and conducting health surveys: A comprehensive guide* (3rd ed.). San Francisco: Jossey-Bass.
- Adler, M., & Ziglio, E. (1996). *Gazing into the oracle: The Delphi method and its application to social policy and public health*. London: Jessica Kingsley Publishers.
- Alkin, M. C. (2004). *Evaluation roots: Tracing theorists’ views and influences*. Thousand Oaks: SAGE.
- Allegrante, J. P., & Livingood, W. C. (2013). IOM report redefines the assessment and value of community-based health prevention. *Health Education & Behavior*, 40(1), 3–5.

- Altman, D. G., Schulz, K. F., Moher, D., Egger, M., Davidoff, F., Elbourne, D., et al. (2001). The revised CONSORT statement for reporting randomized trials: Explanation and elaboration. *Annals of Internal Medicine*, 134(8), 663–694.
- Altman, D. G., Simera, I., Hoey, J., Moher, D., & Schulz, K. (2008). EQUATOR: Reporting guidelines for health research. *Lancet*, 371(9619), 1149–1150.
- Anderson, K. K., Sebaldt, R. J., Lohfeld, L., Karwalajtys, T., Ismaila, A. S., Goeree, R., et al. (2008). Patient views on reminder letters for influenza vaccinations in an older primary care patient population – A mixed methods study. *Canadian Journal of Public Health-Revue Canadienne De Sante Publique*, 99(2), 133–136.
- Andersen, P. A., Buller, D. B., Walkosz, B. J., Maloy, J., Scott, M. D., Cutter, G. R., & Dignan, M. B. (2009). Testing a theory-based health communication program: A replication of Go Sun Smart in outdoor winter recreation. *Journal of Health Communication*, 14(4), 346–365.
- Baker, E. A., Homan, S., Schonhoff, R., & Kreuter, M. (1999). Principles of practice for academic/practice/community research partnerships. *American Journal of Preventive Medicine*, 16(3 Suppl), 86–93.
- Bakken, S., Holzemer, W. L., Brown, M. A., Powell-Cope, G. M., Turner, J. G., Inouye, J., et al. (2000). Relationships between perception of engagement with health care provider and demographic characteristics, health status, and adherence to therapeutic regimen in persons with HIV/AIDS. *AIDS Patient Care and STDs*, 14(4), 189–197.
- Begg, C., Cho, M., Eastwood, S., Horton, R., Moher, D., Olkin, I., et al. (1996). Improving the quality of reporting of randomized trials. The CONSORT statement. *Journal of the American Medical Association*, 276(8), 637–639.
- Benning, A., Ghaleb, M., Suokas, A., Dixon-Woods, M., Dawson, J., Barber, N., et al. (2011). Large scale organisational intervention to improve patient safety in four UK hospitals: Mixed method evaluation. *British Medical Journal*, 342, d195.
- Bonuck, K. A., & Kahn, R. (2002). Prolonged bottle use and its association with iron deficiency anemia and overweight: A preliminary study. *Clinical Pediatrics*, 41(8), 603–607.
- Borawski, E. A., Trapl, E. S., Adams-Tufts, K., Hayman, L. L., Goodwin, M. A., & Lovegreen, L. D. (2009). Taking be proud! Be responsible! To the suburbs: A replication study. *Perspectives on Sexual and Reproductive Health*, 41(1), 12–22.
- Borland, R., Yong, H. H., Wilson, N., Fong, G. T., Hammond, D., Cummings, K. M., Hosking, W., & McNeill, A. (2009). How reactions to cigarette packet health warnings influence quitting: Findings from the ITC Four-Country survey. *Addiction*, 104(4), 669–675.
- Bryson, J. M., Patton, M. Q., & Bowman, R. A. (2011). Working with evaluation stakeholders: A rationale, step-wise approach and toolkit. *Evaluation and Program Planning*, 34(1), 1–12.
- Bullinger, M. (1995). German translation and psychometric testing of the SF-36 Health Survey: Preliminary results from the IQOLA project. *International Quality of Life Assessment. Social Science & Medicine*, 41(10), 1359–1366.
- Burney, P., Luczynska, C., Chinn, S., & Jarvis, D. (1994). The European Community Respiratory Health Survey. *European Respiratory Journal*, 7(5), 954–960.
- Canadian Institutes of Health Research. (2014). *CIHR open access policy*. Retrieved from <http://www.cihr-irsc.gc.ca/e/32005.html>
- Carleton, R. A., Lasater, T. M., Assaf, A. R., Feldman, H. A., & McKinlay, S. (1995). The Pawtucket Heart Health Program: Community changes in cardiovascular risk factors and projected disease risk. *American Journal of Public Health*, 85(6), 777–785.
- Carpenter, S. (2012). Psychology’s bold initiative. *Science*, 335(6076), 1558–1561.
- Centers for Disease Control and Prevention. (2014). *HIV among youth*. Retrieved from http://www.cdc.gov/hiv/risk/age/youth/index.html?s_cid=tw_std0141316
- Centers for Disease Control and Prevention. (2015a). *National Health Interview Survey*. Retrieved from <http://www.cdc.gov/nchs/nhis.htm>
- Centers for Disease Control and Prevention. (2015b). *Behavioral Risk Factor Surveillance System*. Retrieved from: http://www.cdc.gov/brfss/data_documentation/index.htm
- Centers for Disease Control and Prevention. (2015c). *National Health and Nutrition Examination Survey*. Retrieved from <http://www.cdc.gov/nchs/nhanes.htm>
- Chatterji, M. (2004). Evidence on “what works”: An argument for extended-term mixed-method (ETMM) evaluation designs. *Educational Researcher*, 33(9), 3–13.
- Chatterji, M., Green, L. W., & Kumanyika, S. (2001). L.E.A.D.: A prevention framework for evidence gathering and use for the prevention of obesity and other complex health problems. *Health Education & Behavior*, 41(1), 85–99.
- Chatterji, M., Kwon, Y. A., & Sng, C. (2006). Gathering evidence on an after-school supplemental instruction program: Design challenges and early findings in light of NCLB. *Education Policy Analysis Archives*, 14(12), 1–44.
- Clark, V. L. P. (2010). The adoption and practice of mixed methods: US trends in federally funded health-related research. *Qualitative Inquiry*, 16(6), 428–440.
- Cochrane Collaboration. (2015). Retrieved from <http://www.cochrane.org/>
- Cohall, A. T., Dini, S., Senathirajah, Y., Nye, A., Neu, N., Powell, D., et al. (2008). Feasibility of using computer-assisted interviewing to enhance HIV test counseling in community settings. *Public Health Reports*, 123(Suppl 3), 70–77.
- Cohen, D. J., Crabtree, B. F., Etz, R. S., Balasubramanian, B. A., Donahue, K. E., Leviton, L. C., et al. (2008). Fidelity versus flexibility: Translating evidence-based research into practice. *American Journal of Preventive Medicine*, 35(5 Suppl), S381–S389.
- Commission on the Social Determinants of Health. (2008). *Closing the gap in a generation: Health equity through action on the social determinants of health. Final report of the Commission on Social Determinants of Health*. Geneva: World Health Organization.

- Coons, S. J., Alabdulmohsin, S. A., Draugalis, J. R., & Hays, R. D. (1998). Reliability of an Arabic version of the RAND-36 Health Survey and its equivalence to the US-English version. *Medical Care*, *36*(3), 428–432.
- Creswell, J. W., Klassen, A. C., Plano Clark, V. L., & Smith, K. C. for the Office of Behavioral and Social Sciences Research. (2011). *Best practices for mixed methods research in the health sciences*. Retrieved from www.obsr.od.nih.gov/wp-content/uploads/2016/02/Best_Practices_for_Mixed_Methods_Research.pdf
- Curry, L. A., Nembhard, I. M., & Bradley, E. H. (2009). Qualitative and mixed methods provide unique contributions to outcomes research. *Circulation*, *119*(10), 1442–1452.
- de Vries, H., Weijts, W., Dijkstra, M., & Kok, G. (1992). The utilization of qualitative and quantitative data for health education program planning, implementation, and evaluation: A spiral approach. *Health Education & Behavior*, *19*(1), 101–115.
- Dehar, M. A., Casswell, S., & Duignan, P. (1993). Formative and process evaluation of health promotion and disease prevention programs. *Evaluation Review*, *17*(2), 204–220.
- Denny-Smith, T., Bairan, A., & Page, M. C. (2006). A survey of female nursing students' knowledge, health beliefs, perceptions of risk, and risk behaviors regarding human papillomavirus and cervical cancer. *Journal of the American Academy of Nurse Practitioners*, *18*(2), 62–69.
- Des Jarlais, D. C., Lyles, C., Crepaz, N., & Grp, T. (2004). Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: The TREND statement. *American Journal of Public Health*, *94*(3), 361–366.
- Dillman, D. A., Smyth, J. D., & Christian, L. M. (2008). *Internet, mail, and mixed-mode surveys: The tailored design method* (3rd ed.). New York: Wiley.
- Dorfman, L. E., Derish, P. A., & Cohen, J. B. (1992). Hey girlfriend: An evaluation of AIDS prevention among women in the sex industry. *Health Education & Behavior*, *19*(1), 25–40.
- Dubois-Arber, F., Jeannin, A., & Spencer, B. (1999). Long term global evaluation of a national AIDS prevention strategy: The case of Switzerland. *AIDS*, *13*(18), 2571–2582.
- Duke University Health System. (2018). *Duke patient safety center*. Retrieved from: <http://www.dukepatient-safetycenter.com/>
- Edwards, P. K. (1987). Conceptual and methodological issues in evaluating emergent programs. *Evaluation and Program Planning*, *10*(1), 27–33.
- Eidsdottir, S. T., Kristjansson, A. L., Sigfusdottir, I. D., & Allegrante, J. P. (2008). Trends in physical activity and participation in sports clubs among Icelandic adolescents. *European Journal of Public Health*, *18*(3), 289–293.
- Eidsdottir, S. T., Kristjansson, A. L., Sigfusdottir, I. D., Garber, C. E., & Allegrante, J. P. (2010). Trends in body mass index among Icelandic adolescents and young adults from 1992 to 2007. *International Journal of Environmental Research and Public Health*, *7*(5), 2191–2207.
- Eidsdottir, S. T., Kristjansson, A. L., Sigfusdottir, I. D., Garber, C. E., & Allegrante, J. P. (2013). Secular trends in overweight and obesity among Icelandic adolescents: Does socioeconomic status and family structure matter? *Scandinavian Journal of Public Health*, *41*(4), 384–391.
- Eidsdottir, S. T., Kristjansson, A. L., Sigfusdottir, I. D., Garber, C. E., & Allegrante, J. P. (2014). Association between higher BMI and depressive symptoms in Icelandic adolescents: The mediational function of body image. *European Journal of Public Health*, *24*(6), 888–892.
- Epping-Jordan, J. (2004). Research to practice: International dissemination of evidence-based behavioral medicine. *Annals of Behavioral Medicine*, *28*(2), 81–87.
- Evans, R. I., Raines, B. E., & Owen, A. E. (1989). Formative evaluation in school-based health promotion investigations. *Preventive Medicine*, *18*(2), 229–234.
- Farquhar, J. W., Fortmann, S. P., Flora, J. A., Taylor, C. B., Haskell, W. L., Williams, P. T., et al. (1990). Effects of communitywide education on cardiovascular disease risk factors. The Stanford Five-City project. *Journal of the American Medical Association*, *264*(3), 359–365.
- Fenton, K. A., Chinouya, M., Davidson, O., Copas, A., & Team, M. S. (2002). HIV testing and high risk sexual behaviour among London's migrant African communities: A participatory research study. *Sexually Transmitted Infections*, *78*(4), 241–245.
- Fetterman, D. M. (2000). *Foundations of empowerment evaluation*. Thousand Oaks: SAGE.
- Fisher, E. B. (2008). The importance of context in understanding behavior and promoting health. *Annals of Behavioral Medicine*, *35*(1), 3–18.
- Fisher, E. B., Fitzgibbon, M. L., Glasgow, R. E., Haire-Joshu, D., Hayman, L. L., Kaplan, R. M., Nanney, M. S., & Ockene, J. K. (2011). Behavior matters. *American Journal of Preventive Medicine*, *40*(5), e15–e30.
- Flowers, P., Duncan, B., & Knussen, C. (2003). Re-appraising HIV testing: An exploration of the psychosocial costs and benefits associated with learning one's HIV status in a purposive sample of Scottish gay men. *British Journal of Health Psychology*, *8*(2), 179–194.
- Fortmann, S. P., & Varady, A. N. (2000). Effects of a community-wide health education program on cardiovascular disease morbidity and mortality – The Stanford Five-City project. *American Journal of Epidemiology*, *152*(4), 316–323.
- Fowler, F. J., Jr. (2014). *Survey research methods* (5th ed.). Thousand Oaks: SAGE.
- Freudenberg, N. (1982). Health education for social change: A strategy for public health in the US. *International Journal of Health Education*, *24*(3), 138–145.
- Freudenberg, N., & Tsui, E. (2014). Evidence, power, and policy change in community-based participatory research. *American Journal of Public Health*, *104*(1), 11–14.
- Freudenberg, N., Franzosa, E., Chisholm, J., & Libman, K. (2015). New approaches for moving upstream: How state and local health departments can trans-

- form practice to reduce health inequalities. *Health Education & Behavior*, 42(Suppl 1), 46S–56S.
- Fukuhara, S., Bito, S., Green, J., Hsiao, A., & Kurokawa, K. (1998). Translation, adaptation, and validation of the SF-36 Health Survey for use in Japan. *Journal of Clinical Epidemiology*, 51(11), 1037–1044.
- Gabriele, J. M., Carpenter, B. D., Tate, D. F., & Fisher, E. B. (2011). Directive and nondirective e-coach support for weight loss in overweight adults. *Annals of Behavioral Medicine*, 41(2), 252–263.
- Galuska, D. A., Will, J. C., Serdula, M. K., & Ford, E. S. (1999). Are health care professionals advising obese patients to lose weight? *Journal of the American Medical Association*, 282(16), 1576–1578.
- Gandek, B., Ware, J. E., Aaronson, N. K., Alonso, J., Apolone, G., Bjorner, J. B., et al. (1998). Cross-validation of item selection and scoring for the SF-12 Health Survey in nine countries: Results from the IQOLA project. *Journal of Clinical Epidemiology*, 51(11), 1171–1178.
- Gevers, D. W. M., Kremers, S. P. J., de Vries, N. K., & van Assema, P. (2014). Clarifying concepts of food parenting practices. A Delphi study with an application to snacking behavior. *Appetite*, 79, 51–57.
- Gignac, M. A. M., Davis, A. M., Hawker, G., Wright, J. G., Mahomed, N., Fortin, P. R., & Bradley, E. M. (2006). “What do you expect? You’re just getting older”: A comparison of perceived osteoarthritis-related and aging-related health experiences in middle- and older-age adults. *Arthritis and Rheumatism*, 55(6), 905–912.
- Gittelsohn, J., Steckler, A., Johnson, C. C., Pratt, C., Grieser, M., Piclcrel, J., et al. (2006). Formative research in school and community-based health programs and studies: “state of the art” and the TAAG approach. *Health Education & Behavior*, 33(1), 25–39.
- Glasgow, R. E. (2008). What types of evidence are most needed to advance behavioral medicine? *Annals of Behavioral Medicine*, 35(1), 19–25.
- Glasgow, R. E. (2013). What does it mean to be pragmatic? Pragmatic methods, measures, and model to facilitate research translation. *Health Education & Behavior*, 40(3), 257–265.
- Glasgow, R. E., & Emmons, K. M. (2007). How can we increase translations of research into practice? Types of evidence needed. *Annual Review of Public Health*, 28, 413–433.
- Glasgow, R. E., Vogt, T. M., & Boles, S. M. (1999). Evaluating the public health of health promotion interventions: The RE-AIM framework. *American Journal of Public Health*, 89(9), 1322–1327.
- Glasgow, R. E., Eakin, E. G., Fisher, E. B., Bacak, S. J., & Brownson, R. C. (2001). Physician advice and support for physical activity: Results from a national survey. *American Journal of Preventive Medicine*, 21(3), 189–196.
- Glasgow, R. E., Bull, S. S., Gillette, C., Klesges, L. M., & Dzawaltowski, D. A. (2002). Behavior change intervention research in healthcare settings – A review of recent reports with emphasis on external validity. *American Journal of Preventive Medicine*, 23(1), 62–69.
- Glasgow, R. E., Kurz, D., King, D., Dickman, J. M., Faber, A. J., et al. (2010). Outcomes of minimal and moderate support versions of an internet-based diabetes self-management support program. *Journal of General Internal Medicine*, 25(1), 1315–1322.
- Glass, T. A., & McAtee, M. J. (2006). Behavioral science at the crossroads in public health: Extending horizons, envisioning the future. *Social Science & Medicine*, 62(7), 1650–1671.
- Golden, S. D., & Earp, J. L. (2012). Social ecological approaches to individuals and their contexts: Twenty years of health education & behavior health promotion interventions. *Health Education & Behavior*, 39(3), 364–372.
- Gottlieb, N. H., Lovato, C. Y., Weinstein, R., Green, L. W., & Eriksen, M. P. (1992). The implementation of a restrictive worksite smoking policy in a large decentralized organization. *Health Education & Behavior*, 19(1), 77–100.
- Graham, A. L., Papandonatos, G. D., DePue, J. D., Pinto, B. M., Borrelli, B., Neighbors, C. J., et al. (2008). Lifetime characteristics of participants and non-participants in a smoking cessation trial: Implications for external validity and public health impact. *Annals of Behavioral Medicine*, 35(3), 295–307.
- Green, L. W. (2004). Ethics and community-based participatory research: Commentary on Minkler. *Health Education & Behavior*, 31(6), 698–701.
- Green, L. W. (2006). Public health asks of systems science: To advance our evidence-based practice, can you help us get more practice-based evidence? *American Journal of Public Health*, 96(3), 406–409.
- Green, L. W. (2008). Making research relevant: If it is an evidence-based practice, where’s the practice-based evidence? *Family Practice*, 25(Suppl 1), i20–i44.
- Green, L. W. (2014). *The PRECEDE-PROCEED model of health program planning and evaluation*. Bibliography. Published Applications of the Precede Model. Retrieved from <http://www.lgreen.net/precede%20apps/preapps-NEW.htm>
- Green, L. W., & Kreuter, M. (2005). *Health program planning: An educational and ecological approach* (4th ed.). New York: McGraw-Hill.
- Green, L. W., & Mercer, S. L. (2001). Can public health researchers and agencies reconcile the push from funding bodies and the pull from communities? *American Journal of Public Health*, 91(12), 1926–1929.
- Greene, J. C. (1987). Stakeholder participation in evaluation design: Is it worth the effort? *Evaluation and Program Planning*, 10(4), 379–394.
- Greenhalgh, T., Helman, C., & Chowdhury, A. M. (1998). Health beliefs and folk models of diabetes in British Bangladeshis: A qualitative study. *British Medical Journal*, 316(7136), 978–983.
- Greenwald, A. G. (1976). Within-subjects designs: To use or not to use? *Psychological Bulletin*, 83(2), 314–320.
- Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park: SAGE.
- Hallfors, D., Cho, H., Sanchez, V., Khatapoush, S., Kim, H. M., & Bauer, D. (2006). Efficacy vs effectiveness trial results of an indicated “model” substance abuse

- program: Implications for public health. *American Journal of Public Health*, 96(12), 2254–2259.
- Hammond, D., Fong, G., McDonald, P. W., Brown, S., & Cameron, R. (2004). Graphic Canadian cigarette warning labels and adverse outcomes: Evidence from Canadian smokers. *American Journal of Public Health*, 94(8), 1442–1445.
- Hanson, D. W., Finch, C. F., Allegrante, J. P., & Sleet, D. (2012). Closing the gap between injury prevention research and community safety promotion practice: Revisiting the public health model. *Public Health Reports*, 127(2), 147–155.
- HealthIT.gov. *How do I use a rapid-cycle improvement strategy?* Retrieved from <http://www.healthit.gov/providers-professionals/faqs/how-do-i-use-rapid-cycle-improvement-strategy>
- Helft, P. R., Hlubocky, F., & Daugherty, C. K. (2003). American oncologists' views of internet use by cancer patients: A mail survey of American Society of Clinical Oncology members. *Journal of Clinical Oncology*, 21(5), 942–947.
- Higgins, D. L., O'Reilly, K., Tashima, N., Crain, C., Beeker, C., Goldbaum, G., et al. (1996). Using formative research to lay the foundation for community level HIV prevention efforts: An example from the AIDS community demonstration projects. *Public Health Reports*, 111(Suppl 1), 28–35.
- Higgins, J. P., Altman, D. G., Gøtzsche, P. C., Jüni, P., Moher, D., Oxman, A. D., Savovic, J., Schulz, K. F., Weeks, L., Sterne, J. A., & Cochrane Bias Methods Group; Cochrane Statistical Methods Group. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *British Medical Journal*, 343, d5928.
- Homer, J., Milstein, B., Wile, K., Pratibhu, P., Farris, R., & Orenstein, D. (2008). Modeling the local dynamics of cardiovascular health: Risk factors, context, and capacity. *Preventing Chronic Disease*, 5(2), A63.
- Homer, J., Milstein, B., Wile, K., Trogon, J., Huang, P., Labarthe, D., & Orenstein, D. (2010). Simulating and evaluating local interventions to improve cardiovascular health. *Preventing Chronic Disease*, 7(1), A18.
- Honeycutt, A. A., Wile, K., Dove, C., Hawkins, J., & Orenstein, D. (2015). Strategic planning for chronic disease prevention in rural America: Looking through a PRISM lens. *Journal of Public Health Management and Practice*, 21(4), 392–329.
- Hyden, C., Kahn, R., & Bonuck, K. (2013). Bottle-weaning intervention tools: The “how” and “why” of a WIC-based educational flipchart, parent brochure, and website. *Health Promotion Practice*, 14(1), 75–80.
- IOM (Institute of Medicine). (2011). *Finding what works in health care: Standards for systematic reviews*. Washington, DC: The National Academies Press.
- IOM (Institute of Medicine). (2012). *An integrated framework for assessing the value of community-based prevention*. Washington, DC: The National Academies Press.
- Isaac, S., & Michael, W. B. (1995). *Handbook in research and evaluation: A collection of principles, methods, and strategies useful in the planning, design, and evaluation of studies in education and the behavioral sciences* (3rd ed.). San Diego: EdITS.
- Israel, B. A., Schulz, A. J., Parker, E. A., & Becker, A. B. (1998). Review of community based research: Assessing partnership approaches to improve public health. *Annual Review of Public Health*, 19(1), 173–202.
- Israel, B. A., Schulz, A. J., Parker, E. A., Becker, A. B., & Community-Campus Partnerships for Health. (2001). Community-based participatory research: Policy recommendations for promoting a partnership approach in health research. *Education for Health*, 14(2), 182–197.
- Jacobs, C. L., Dehue, F., Vollink, T., & Lechner, L. (2014). Determinants of adolescents' ineffective and improved coping with cyberbullying: A Delphi study. *Journal of Adolescence*, 37(4), 373–385.
- Jarvandi, S., Davidson, N. O., Jeffe, D. B., & Schootman, M. (2012). Influence of lifestyle factors on inflammation in men and women with type 2 diabetes: Results from the National Health and Nutrition Examination Survey, 1999–2004. *Annals of Behavioral Medicine*, 44(3), 399–407.
- Jasny, B. R., Chin, G., Chong, L., & Vignieri, S. (2011). Data replication & reproducibility. Again, and again, and again Introduction. *Science*, 334(6060), 1225.
- Johnson, B., & Christensen, L. B. (2004). *Educational research: Quantitative, qualitative and mixed approaches* (2nd ed.). Boston: Allyn & Bacon.
- Johnson, J. A., Nowatzi, T. E., & Coons, S. J. (1996). Health-related quality of life of diabetic Pima Indians. *Medical Care*, 34(2), 97–102.
- Katz, D. L., Murimi, M., Gonzalez, A., Njike, V., & Green, L. W. (2011). From controlled trial to community adoption: The multisite translational community trial. *American Journal of Public Health*, 101(8), e17–e27.
- Kelley, K., Clark, B., Brown, V., & Sitzia, J. (2003). Good practice in the conduct and reporting of survey research. *International Journal for Quality in Health Care*, 15(3), 261–266.
- Kenny, S. J., Smith, P. J., Goldschmid, M. G., Newman, J. M., & Herman, W. H. (1993). Survey of physician practice behaviors related to diabetes mellitus in the U.S.: Physician adherence to consensus recommendations. *Diabetes Care*, 16(11), 1507–1510.
- Kim, Y. S., Park, Y. S., Allegrante, J. P., Marks, R., Ok, H., Cho, K., & Garber, C. E. (2012). Relationship between physical activity and general mental health. *Preventive Medicine*, 55(5), 458–463.
- Kisker, E. E., & Brown, R. S. (1997). Nonexperimental designs and program evaluation. *Children and Youth Services Review*, 19(7), 541–566.
- Kitzinger, J. (1994). The methodology of focus groups: The importance of interaction between research participants. *Sociology of Health & Illness*, 16(1), 103–121.
- Knickman, J., & Jellinek, P. (1997). Epilogue: Four lessons from evaluating controversial programs. *Children and Youth Services Review*, 19(7), 607–614.
- Knuth, A. G., Malta, D. C., Dumith, S. C., Pereira, C. A., Morais-Neto, O. L., Temporao, J. G., et al. (2011). Practice of physical activity and sedentarism among Brazilians: Results of the National Household

- Sample Survey-2008. *Ciencia & Saude Coletiva*, 16(9), 3697–3705.
- Kobeissi, L., Nakkash, R., Ghantous, Z., Abou Saad, M., & Yassin, N. (2011). Evaluating a community based participatory approach to research with disadvantaged women in the southern suburbs of Beirut. *Journal of Community Health*, 36(5), 741–747.
- Kolata, G. (2013, November 25). Bumps in the road to new cholesterol guidelines. *The New York Times*. Retrieved from http://www.nytimes.com/2013/11/26/health/heart-and-stroke-study-hit-by-a-wave-of-criticism.html?pagewanted=all&_r=0
- Korelitz, J. J., Fernandez, A. A., Uyeda, V. J., Spivey, G. H., Browdy, B. L., & Schmidt, R. T. (1993). Health habits and risk factors among truck drivers visiting a health booth during a trucker trade show. *American Journal of Health Promotion*, 8(2), 117–123.
- Kreuger, R. A., & Casey, M. A. (2015). *Focus groups: A practical guide for applied research* (5th ed.). Thousand Oaks: SAGE.
- Kristjansson, A. L., James, J. E., Allegrante, J. P., Sigfusdottir, I. D., & Helgason, A. R. (2010). Adolescent substance use, parental monitoring, and leisure-time activities: 12-year outcomes of primary prevention in Iceland. *Preventive Medicine*, 51(2), 168–171.
- Kristjansson, A. L., Sigfusson, J., Sigfusdottir, I. D., & Allegrante, J. P. (2013). Data collection procedures for school-based surveys among adolescents: The Youth in Europe Study. *Journal of School Health*, 83(9), 662–667.
- Kristjansson, A. L., Sigfusdottir, I. D., Sigfusson, J., & Allegrante, J. P. (2014). Self-generated identification codes in longitudinal prevention research with adolescents: A pilot study of matched and unmatched subjects. *Prevention Science*, 15(2), 205–212.
- Kurth, B. M., Kamtsiuris, P., Holling, H., Schlaud, M., Dolle, R., Ellert, U., et al. (2008). The challenge of comprehensively mapping children's health in a nation-wide health survey: Design of the German KiGGS-Study. *BMC Public Health*, 8, 196.
- Langley, C., Nolan, K., Nolan, T., Norman, C., & Provost, L. (1996). *The improvement guide: A practical approach to improving organizational performance*. San Francisco: Jossey-Bass.
- Lantz, P. M., House, J. S., Lepkowski, J. M., Williams, D. R., Mero, R. P., & Chen, J. M. (1998). Socioeconomic factors, health behaviors, and mortality: Results from a nationally representative prospective study of US adults. *Journal of the American Medical Association*, 279(21), 1703–1708.
- Lear, J. G., Gleicher, H. B., St Germaine, A., & Porter, P. J. (1991). Reorganizing health care for adolescents: The experience of the school-based adolescent health care program. *Journal of Adolescent Health*, 12(6), 450–458.
- Levy, R. I. (1982). The National Heart, Lung, and Blood Institute overview 1980. The director's report to the NHLBI Advisory Council. *Circulation*, 65(2), 217–225.
- Li, L., Wang, H. M., & Shen, Y. (2003). Chinese SF-36 Health Survey: Translation, cultural adaptation, validation, and normalisation. *Journal of Epidemiology and Community Health*, 57(4), 259–263.
- Li, Y., Ehiri, J., Hu, D., Zhang, Y., Wang, Q., Zhang, S., & Cao, J. (2014). Framework of behavioral indicators for outcome evaluation of TB health promotion: A Delphi study of TB suspects and Tb patients. *BMC Infectious Diseases*, 14, 268.
- Lieberman, L., Golden, S. D., & Earp, J. A. (2013). Structural approaches to health promotion: What do we need to know about policy and environmental change? *Health Education & Behavior*, 40(5), 520–525.
- Lindsay, A. C. (2002). Integrating quantitative and qualitative methods to assess the impact of child survival programs in developing countries: The case of a program evaluation in Ceara, Northeast Brazil. *Health Education & Behavior*, 29(5), 570–584.
- Linstone, H. A., & Turoff, M. (Eds.). (1975). *The Delphi method: Techniques and applications*. Reading: Addison-Wesley Publishing Company/Advanced Book Program.
- Linstone, H. A., & Turoff, M. (2011). Delphi: A brief look backward and forward. *Technological Forecasting and Social Change*, 78(9), 1712–1719.
- Livingood, W. C., Allegrante, J. P., Airhihenbuwa, C. O., Clark, N. M., Windsor, R. C., Zimmerman, M. A., & Green, L. W. (2011). Applied social and behavioral science to address complex health problems. *American Journal of Preventive Medicine*, 41(5), 525–531.
- Livingood, W. C., Allegrante, J. P., & Green, L. W. (2016). Culture change from tobacco accommodation to intolerance: Time to connect the dots. *Health Education & Behaviour*, 43(2), 133–138.
- Luepker, R. V., Murray, D. M., Jacobs, D. R., Jr., Mittelman, M. B., Bracht, N., Carlaw, R., et al. (1994). Community education for cardiovascular disease prevention: Risk factor changes in the Minnesota Heart Health Program. *American Journal of Public Health*, 84(9), 1383–1393.
- Luostarinen, T., Hakulinen, T., & Pukkala, E. (1995). Cancer risk following a community-based programme to prevent cardiovascular diseases. *International Journal of Epidemiology*, 24(6), 1094–1099.
- Macaulay, A. C., Jagosh, J., Sellar, R., Henderson, J., Cargo, M., Greenhalgh, T., et al. (2011). Assessing the benefits of participatory research: A rationale for a realist review. *Global Health Promotion*, 18(2), 45–48.
- Manca, D. P., Aubrey-Bassler, K., Kandola, K., Aquilar, C., Campbell-Scherer, D., Sopcak, N., et al. (2014). Implementing and evaluating a program to facilitate chronic disease prevention and screening in primary care: A mixed methods program evaluation. *Implementation Science*, 9, 135.
- Marcoux, M. F., Sallis, J. F., McKenzie, T. L., Marshall, S., Armstrong, C. A., & Goggin, K. J. (1999). Process evalu-

- ation of a physical activity self-management program for children: SPARK. *Psychology & Health*, 14(4), 659–677.
- McBurney, D. H., & White, T. L. (2010). *Research methods* (8th ed.). Belmont: Thomson/Wadsworth.
- McGraw, S. A., McKinlay, S. M., McClements, L., Lasater, T. M., Assaf, A., & Carleton, R. A. (1989). Methods in program evaluation: The process evaluation system of the Pawtucket Heart Health Program. *Evaluation Review*, 13(5), 459–483.
- Meadows, G., Harvey, C., Fossey, E., & Burgess, P. (2000). Assessing perceived need for mental health care in a community survey: Development of the Perceived Need for Care Questionnaire (PNCQ). *Social Psychiatry and Psychiatric Epidemiology*, 35(9), 427–435.
- Mermelstein, R. (1999). Explanations of ethnic and gender differences in youth smoking: A multi-site, qualitative investigation. *Nicotine & Tobacco Research*, 1(Suppl 1), S91–S98.
- Minkler, M., Blackwell, A. G., Thompson, M., & Tamir, H. (2003). Community-based participatory research: Implications for public health funding. *American Journal of Public Health*, 93(8), 1210–1213.
- Minkler, M., Vasquez, V. B., Warner, J. R., Stussey, H., & Facente, S. (2006). Sowing the seeds for sustainable change: A community-based participatory research partnership for health promotion in Indiana, USA and its aftermath. *Health Promotion International*, 21(4), 293–300.
- Minkler, M., Lee, P. T., Tom, A., Chang, C., Morales, A., Liu, S. S., et al. (2010). Using community-based participatory research to design and initiate a study on immigrant worker health and safety in San Francisco's Chinatown restaurants. *American Journal of Industrial Medicine*, 53(4), 361–371.
- Moen, R. D., Nolan, T. W., & Provost, L. P. (1999). *Quality improvement through planned experimentation*. New York: McGraw-Hill.
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *British Medical Journal*, 339, b2535.
- Montgomery, P., Grant, S., Hopewell, S., Macdonald, G., Moher, D., Michie, S., & Mayo-Wilson, E. (2013). Protocol for CONSORT-SPI: An extension for social and psychological interventions. *Implementation Science*, 8, 99.
- Morgan, D. L. (1996). Focus groups. *Annual Review of Sociology*, 22(1), 129–152.
- Narang, R. (2011). Determining quality of public health care services in rural India. *Clinical Governance: An International Journal*, 16(1), 35–49.
- National Institutes of Health. (2014). *Public access policy*. Retrieved from <http://publicaccess.nih.gov/>
- Nature. (2006). Overview: *Nature's* peer review trial. Retrieved from <http://www.nature.com/nature/peer-review/debate/nature05535.html>
- NCI (National Cancer Institute). (2015). *Health information national trends survey*. Retrieved from <http://hints.cancer.gov/>
- Newhouse, J. P. (1982). A summary of the Rand Health Insurance Study. *Annals of the New York Academy of Sciences*, 387, 111–114.
- Ngo-Metzger, Q., Sorkin, D. H., Mangione, C. M., Gandek, B., & Hays, R. D. (2008). Evaluating the SF-36 Health Survey (Version 2) in older Vietnamese Americans. *Journal of Aging and Health*, 20(4), 420–436.
- Nicklas, T. A., & O'Neil, C. E. (2000). Process of conducting a 5-a-day intervention with high school students: Gimme 5 (Louisiana). *Health Education & Behavior*, 27(2), 201–212.
- O'Brien, K. (1993). Using focus groups to develop health surveys: An example from research on social relationships and AIDS-preventive behavior. *Health Education & Behavior*, 20(3), 361–372.
- O'Brien, B. C., Harris, I. B., Beckman, T. J., Reed, D. A., & Cook, D. A. (2014). Standards for reporting qualitative research: A synthesis of recommendations. *Academic Medicine*, 89(9), 1245–1251.
- Oldenburg, B. F., Sallis, J. F., EFrench, M. L., & Owen, N. (1999). Health promotion research and the diffusion and institutionalization of interventions. *Health Education Research*, 14, 121–130.
- Owen, L., & Penna, G. (2001). *Smoking and pregnancy: A survey of knowledge attitudes and behaviour 1992–1999*. London: National Institute for Health and Clinical Excellence.
- Parker, R., Baker, D., Williams, M., & Nurss, J. (1995). The test of functional health literacy in adults. *Journal of General Internal Medicine*, 10(10), 537–541.
- Patton, M. Q. (1984). An alternative evaluation approach for the problem-solving training program: A utilization-focused evaluation process. *Evaluation and Program Planning*, 7(2), 189–192.
- Patton, M. Q. (1987). *How to use qualitative methods in evaluation*. Newbury Park: SAGE.
- Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks: SAGE.
- Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). Thousand Oaks: SAGE.
- Patton, M. Q. (2012). *Essentials of utilization-focused evaluation*. Thousand Oaks: SAGE.
- Pazoki, R., Nabipour, I., Seyednezami, N., & Imami, S. R. (2007). Effects of a community-based healthy heart program on increasing healthy women's physical activity: A randomized controlled trial guided by Community-based Participatory Research (CBPR). *BMC Public Health*, 7, 216.
- Perneger, T. V., Leplège, A., Etter, J.-F., & Rougemont, A. (1995). Validation of a French-language version of the MOS 36-Item Short Form Health Survey (SF-36) in young healthy adults. *Journal of Clinical Epidemiology*, 48(8), 1051–1060.
- Persson, L.-O., Karlsson, J., Bengtsson, C., Steen, B., & Sullivan, M. (1998). The Swedish SF-36 Health Survey II. Evaluation of clinical validity: Results from population studies of elderly and women in Gothenburg. *Journal of Clinical Epidemiology*, 51(11), 1095–1103.

- Peterson, J. C., Charlson, M. E., Hoffman, Z., Wells, M. T., Wong, S. C., Hollenberg, J. P., et al. (2012). A randomized controlled trial of positive-affect induction to promote physical activity after percutaneous coronary intervention. *Archives of Internal Medicine*, *172*(4), 329–336.
- Peterson, J. C., Czajkowski, S., Charlson, M. E., Link, A. R., Wells, M. T., Isen, A. M., et al. (2013). Translating basic behavioral and social science research to clinical application: The EVOLVE mixed methods approach. *Journal of Consulting and Clinical Psychology*, *81*(2), 217–230.
- Peterson, J. C., Link, A. R., Jobe, J. B., Winston, G. J., Marina Limasiewski, E., & Allegrante, J. P. (2014). Developing self-management education in coronary artery disease. *Heart & Lung: The Journal of Critical Care*, *43*(2), 133–139.
- Petrie, K. J., Cameron, L. D., Ellis, C. J., Buick, D., & Weinman, J. (2002). Changing illness perceptions after myocardial infarction: An early intervention randomized controlled trial. *Psychosomatic Medicine*, *64*(4), 580–586.
- Pinto, R. M., & McKay, M. M. (2006). A mixed-method analysis of African-American women's attendance at an HIV prevention intervention. *Journal of Community Psychology*, *34*(5), 601–616.
- Pole, M., Einarson, A., Paireadeau, N., Einarson, T., & Koren, G. (2000). Drug labeling and risk perceptions of teratogenicity: A survey of pregnant Canadian women and their health professionals. *Journal of Clinical Pharmacology*, *40*(6), 573–577.
- Polit, D. F., & Beck, C. T. (2008). *Nursing research: Generating and assessing evidence for nursing practice* (8th ed.). Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins.
- Posavac, E. J. (2011). *Program evaluation: Methods and case studies* (8th ed.). Boston: Prentice Hall.
- Puska, P., Nissinen, A., Tuomilehto, J., Salonen, J. T., Koskela, K., McAlister, A., et al. (1985). The community-based strategy to prevent coronary heart disease: Conclusions from the ten years of the North Karelia project. *Annual Review of Public Health*, *6*, 147–193.
- Puska, P., Vartiainen, E., Tuomilehto, J., Salomaa, V., & Nissinen, A. (1998). Changes in premature deaths in Finland: Successful long-term prevention of cardiovascular diseases. *Bulletin of the World Health Organization*, *76*(4), 419–425.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks: SAGE.
- Rea, L. M., & Parker, R. M. (2014). *Designing and conducting survey research: A comprehensive guide* (4th ed.). San Francisco: Jossey-Bass.
- Robinson, N. (1999). The use of focus group methodology with selected examples from sexual health research. *Journal of Advanced Nursing*, *29*(4), 905–913.
- Rou, K. M., Wu, Z. Y., Sullivan, S. G., Li, F., Guan, J. H., Xu, C., et al. (2007). A five-city trial of a behavioural intervention to reduce sexually transmitted disease/HIV risk among sex workers in China. *AIDS*, *21*(Suppl 8), S95–S101.
- Sandelowski, M. (1995). Sample size in qualitative research. *Research in Nursing & Health*, *18*(2), 179–183.
- Santos, I., Victora, C. G., Martines, J., Gonçalves, H., Gigante, D. P., Valle, N. J., & Pelto, G. (2001). Nutrition counseling increases weight gain among Brazilian children. *The Journal of Nutrition*, *131*(11), 2866–2873.
- Schneider, S., Maul, H., Freerksen, N., & Potschke-Langer, M. (2008). Who smokes during pregnancy? An analysis of the German Perinatal Quality Survey 2005. *Public Health*, *122*(11), 1210–1216.
- Schumacher, K. L., Koresawa, S., West, C., Dodd, M., Paul, S. M., Tripathy, D., et al. (2005). Qualitative research contribution to a randomized clinical trial. *Research in Nursing & Health*, *28*(3), 268–280.
- Secginli, S., & Nahcivan, N. O. (2006). Factors associated with breast cancer screening behaviours in a sample of Turkish women: A questionnaire survey. *International Journal of Nursing Studies*, *43*(2), 161–171.
- Seifer, S. (2006). Building and sustaining community-institutional partnerships for prevention research: Findings from a national collaborative. *Journal of Urban Health*, *84*(3), 461–474.
- Shain, R. N., Piper, J. M., Newton, E. R., Perdue, S. T., Ramos, R., Champion, J. D., & Guerra, F. A. (1999). A randomized, controlled trial of a behavioral intervention to prevent sexually transmitted disease among minority women. *New England Journal of Medicine*, *340*(2), 93–100.
- Sigfusdottir, I. D., Kristjansson, A. L., Thorlindsson, T., & Allegrante, J. P. (2008). Trends in prevalence of substance use among Icelandic adolescents, 1995–2006. *Substance Abuse Treatment, Prevention, and Policy*, *3*, 12.
- Sigfusdottir, I. D., Thorlindsson, T., Kristjansson, A. L., Roe, K. M., & Allegrante, J. P. (2009). Substance use prevention for adolescents: The Icelandic Model. *Health Promotion International*, *24*(1), 16–25.
- Sigfusdottir, I. D., Kristjansson, A. L., Gudmundsdottir, M. L., & Allegrante, J. P. (2011). Substance use prevention through school and community-based health promotion: A transdisciplinary approach from Iceland. *Global Health Promotion*, *18*(3), 23–26.
- Singh, L., Sharma, A., Kumar, M., & Shinwari, S. (2012). Public health care in Afghanistan: An investigation in suboptimal utilization of facilities. *Health*, *4*(10), 794–801.
- Sirois, F. M. (2007). “I’ll look after my health, later”: A replication and extension of the procrastination-health model with community-dwelling adults. *Personality and Individual Differences*, *43*(1), 15–26.
- Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). Thousand Oaks: SAGE.
- Stangor, C. (2007). *Research methods for the behavioral sciences* (3rd ed.). Boston: Houghton Mifflin Co.

- Stanwood, N. L., Garrett, J. M., & Konrad, T. R. (2002). Obstetrician-gynecologists and the intrauterine device: A survey of attitudes and practice. *Obstetrics & Gynecology*, *99*(2), 275–280.
- Steckler, A., & McLeroy, K. R. (2008). The importance of external validity. *American Journal of Public Health*, *98*(1), 9–10.
- Steckler, A., McLeroy, K. R., Goodman, R. M., Bird, S. T., & McCormick, L. (1992). Toward integrating qualitative and quantitative methods: An introduction. *Health Education Quarterly*, *19*(1), 1–8.
- Stewart, K. E., Wright, P. B., Sims, D., Tyner, K. R., & Montgomery, B. E. E. (2012). The “translators”: Engaging former drug users as key research staff to design and implement a risk reduction program for rural cocaine users. *Substance Use & Misuse*, *47*(5), 547–554.
- Story, M. T., Neumark-Sztainer, D. R., Sherwood, N. E., Holt, K., Sofka, D., Trowbridge, F. L., & Barlow, S. E. (2002). Management of child and adolescent obesity: Attitudes, barriers, skills, and training needs among health care professionals. *Pediatrics*, *110*(1), 210–214.
- Tellado, G. S. (1984). An evaluation case: The implementation and evaluation of a problem-solving training program for adolescents. *Evaluation and Program Planning*, *7*(2), 179–188.
- Tews, D. S., Heany, J., Jones, J., VanDerMoere, R., & Madamala, K. (2012). *Embracing quality in public health: A practitioner's quality improvement guidebook*. Retrieved from https://www.mphiaccredandqi.org/wp-content/uploads/2013/12/2012_02_28_Guidebook_web_v2.pdf
- Thorisdottir, I. E., Kristjansson, A. L., Sigfustottir, I. D., & Allegrante, J. P. (2012). The landscape of overweight and obesity in Icelandic adolescents: Geographic variation in body-mass index between 2000 and 2009. *Journal of Community Health*, *37*(1), 234–241.
- Thorpe, K. E., & Yang, Z. (2011). Enrolling people with prediabetes ages 60–64 in a proven weight loss program could save Medicare \$7 billion or more. *Health Affairs*, *30*(9), 1673–1679.
- Tsugane, S., & Sobue, T. (2001). Baseline survey of JPHC study—Design and participation rate. Japan Public Health Center-based Prospective Study on Cancer and Cardiovascular Diseases. *Journal of Epidemiology*, *11*(6 Suppl), S24–S29.
- U.S. Preventive Services Task Force. Retrieved from www.uspreventiveservicestaskforce.org
- UK Research and Innovation (2018). *Open access*. Retrieved from <https://www.ukri.org/funding/information-for-award-holders/open-access/>
- Van Rossem, R., & Meekers, D. (2000). An evaluation of the effectiveness of targeted social marketing to promote adolescent and young adult reproductive health in Cameroon. *AIDS Education and Prevention*, *12*(5), 383–404.
- Vartiainen, E., Puska, P., Jousilahti, P., Korhonen, H. J., Tuomilehto, J., & Nissinen, A. (1994). Twenty-year trends in coronary risk factors in North Karelia and in other areas of Finland. *International Journal of Epidemiology*, *23*(3), 495–504.
- Victora, C. G., Habicht, J.-P., & Bryce, J. (2004). Evidence-based public health: Moving beyond randomized trials. *American Journal of Public Health*, *94*(3), 400–405.
- Viswanathan, M., Ammerman, A., Eng, E., Gartlehner, G., Lohr, K. N., Griffith, D., et al. (2004). *Community based participatory research: Assessing the evidence* (AHRQ Publication No. 04–E022-2). Rockville: U.S. Department of Health and Human Services/Agency for Healthcare Research and Quality. Retrieved from <http://archive.ahrq.gov/downloads/pub/evidence/pdf/cbpr/cbpr.pdf>
- Wallerstein, N. B., & Duran, B. (2006). Using community-based participatory research to address health disparities. *Health Promotion Practice*, *7*(3), 312–323.
- Wallerstein, N. B., & Duran, B. (2010). Community-based participatory research contributions to intervention research: The intersection of science and practice to improve health equity. *American Journal of Public Health*, *100*(Suppl 1), S40–S46.
- Ware, J., Kosinski, M., & Keller, S. D. (1996). A 12-Item Short-Form Health Survey: Construction of scales and preliminary tests of reliability and validity. *Medical Care*, *34*(3), 220–233.
- Warnecke, R. B., Johnson, T. P., Chavez, N., Sudman, S., O'Rourke, D. P., Lacey, L., & Horm, J. (1997). Improving question wording in surveys of culturally diverse populations. *Annals of Epidemiology*, *7*(5), 334–342.
- Weiss, C. H. (1972). *Evaluation research: Methods for assessing program effectiveness*. Englewood Cliffs: Prentice-Hall.
- Whitby, M., McLaws, M.-L., Slater, K., Tong, E., & Johnson, B. (2008). Three successful interventions in health care workers that improve compliance with hand hygiene: Is sustained replication possible? *American Journal of Infection Control*, *36*(5), 349–355.
- White, K. P., Speechley, M., Harth, M., & Ostbye, T. (1995). Fibromyalgia in rheumatology practice: A survey of Canadian rheumatologists. *Journal of Rheumatology*, *22*(4), 722–726.
- Williams, A. M. (2010). Evaluating Canada's compassionate care benefit using a utilization-focused evaluation framework: Successful strategies and prerequisite conditions. *Evaluation and Program Planning*, *33*(2), 91–97.
- Wood, D. L., Valdez, R. B., Hayashi, T., & Shen, A. (1990). Health of homeless children and housed, poor children. *Pediatrics*, *86*(6), 858–866.
- Yang, W., Lu, J., Weng, J., Jia, W., Ji, L., Xiao, J., He, J., & China National Diabetes and Metabolic Disorders Study Group. (2010). Prevalence of diabetes among men and women in China. *New England Journal of Medicine*, *362*(12), 1090–1101.
- Yong, E. (2012). Replication studies: Bad copy. *Nature*. Retrieved from <http://www.nature.com/news/replication-studies-bad-copy-1.10634>
- Yong, H., Borland, R., Thrasher, J. F., Thompson, M. E., Nagelhout, G. E., Fong, G. T., Hammond, D., & Cummings, K. M. (2014). Mediation pathways of

- the impact of cigarette warning labels on quit attempts. *Health Psychology, 33*(11), 1410–1420.
- Young, S. D., Cumberland, W. G., Nianogo, R., Menacho, L. A., Galea, J. T., & Coates, T. (2015). The HOPE social media intervention for global HIV prevention in Peru: A cluster randomised controlled trial. *The Lancet HIV, 2*(1), e27–e32.
- Zapka, J. G., Fletcher, K. E., Ma, Y., & Pbert, L. (1997). Physicians and smoking cessation. Development of survey measures. *Evaluation & the Health Professions, 20*(4), 407–427.
- Zapka, J. G., Pbert, L., Stoddard, A. M., Ockene, J. K., Goins, K. V., & Bonollo, D. (2000). Smoking cessation counseling with pregnant and postpartum women: A survey of community health center providers. *American Journal of Public Health, 90*(1), 78–84.