

This book is out of print. For current reviews, visit www.thecommunityguide.org

Part III

Methodological Background

This book is out of print. For current reviews, visit www.thecommunityguide.org

Chapter 10

Methods Used for Reviewing Evidence and Linking Evidence to Recommendations

Establishing an explicit process for reviewing evidence and linking evidence to recommendations serves to enhance the credibility of those recommendations. This chapter discusses the established methods and process by which evidence is identified, assessed, summarized, and linked to recommendations in the *Community Guide*. Here, we outline the steps in our process, discuss a few updates and lessons learned, and briefly discuss the rationale for our approach.

This chapter is not a manual for conducting systematic reviews, although detailed information on *Community Guide* methods can be found elsewhere.^{1,2} The aim of this chapter is to help users to better understand our methods of developing reviews and recommendations and, therefore, to enhance the credibility of *Community Guide* recommendations. These methods are not the only possible approach, nor are they presented as the best conceivable approach. They reflect various real-life constraints on time, personnel, and other resources that affect the conduct of systematic reviews. Nonetheless, with them, we have generated a large number of credible and transparent findings and recommendations across diverse areas of public health. These *Community Guide* methods have evolved over time and will continue to adapt to enhancements in systematic review methodology, in translating evidence into recommendations, and in meeting user needs.

The methods and process of *Community Guide* reviews are overseen and directed by the independent, nonfederal Task Force on Community Preventive Services, which also makes the evidence-based recommendations. The Task Force's independence helps to insulate the scientific process from real or perceived political pressure. The diversity of expertise on the Task Force well qualifies its members to make the inevitable judgment calls in conducting a review (e.g., Which outcomes should be considered to represent success? How much diversity can be tolerated in a group of related but not identical interventions? What constitutes *consistency* of results?). Task Force members are also well suited to making the inevitable trade-offs in making recommendations (e.g., How much benefit is required to outweigh occasional harms?).

Once a general topic has been selected by the Task Force, the first steps in conducting a review are:

- Set priorities for topics to review
- Convene a systematic review development team

After selecting topics, the systematic review development team:

- Develops a conceptual model for each topic and intervention and specifies the outcomes that will be used to determine if the interventions are effective (that is, the extent to which they actually achieve important goals in promoting health or reducing disease, injury, and impairment).
- Identifies and defines priority interventions.
 - We have defined an *intervention* to be any activity (or group of related activities) intended to prevent disease or promote health in a group of people.
- Systematically searches for and retrieves scientific literature that may provide evidence of the effectiveness of the selected interventions.
- Rates the quality of the studies (both in terms of study design and execution).
- Summarizes the size of any observed effects and the strength of the body of evidence for each intervention across the studies according to the number of studies, their quality, and the size and consistency of effect among the studies. In this step we also consider whether, among all of the available studies, any common threats to validity exist that might reduce confidence in the results.
- Translates the strength of evidence into a recommendation.

The steps are described in more detail in this chapter.

METHODS AND PROCESS FOR DEVELOPING *COMMUNITY GUIDE* REVIEWS AND RECOMMENDATIONS

Select Topics

The Task Force chooses broad topics for review on the basis of the public health burden of the problem; how preventable it is; how it relates to other public health initiatives; and the current level of research and practice activity in public health, clinical, and other settings.³ The agenda-setting process incorporates input from interested others. For example, the first round of priority setting by the Task Force benefited from information provided by subject matter experts from the Centers for Disease Control and Prevention (CDC) and elsewhere in the Department of Health and Human Services.³

Convene a Team

The systematic review development team (the team) of 6–12 people meets frequently to direct the review, and 10–30 additional consultants provide

opinions and expertise as needed. Generalists on the team (including staff, Task Force members, and methodologists) ensure that diverse viewpoints are reflected, that reviews are conducted and communicated in a consistent manner, that rigorous review methods are applied, and that the results can be understood by generalist audiences. Subject matter experts (e.g., experts in the substantive scientific area, experts in the practical implementation of programs and policies, or experts in the particular populations or settings involved) ensure that the questions are relevant to practice, that the information is complete and accurate, and that the reviews and recommendations are conceptually sound. The independence of the Task Force is complemented by the involvement of staff of federal, state, and local agencies, which may have opportunities to act on the results. All of the various collaborators provide valuable and complementary contributions that help to enhance the quality of the reviews and the credibility of the findings, as well as helping to assure that results are useful and used.

Develop a Conceptual Model

We organize our reviews using conceptual models—easy-to-understand diagrams that describe relationships between causes of public health problems (determinants), interventions, intermediate outcomes (such as changes in behavior), and health outcomes. Rather than strictly adhering to one theory, we have used various theories and other conceptual background to inform the development of our conceptual models and to reflect thinking that is relevant to a broad spectrum of public health topics, interventions, and approaches.

We use two types of conceptual models at different stages of the review process. We develop a *logic framework* (Figure 10–1 is one example of a logic framework) early in the systematic review process to illustrate the broad public health context in which a set of interventions might act. Logic frameworks show the “big-picture” relationships among social, environmental, and biological determinants, outcomes, strategic points for action, and interventions. By providing an easy-to-understand visual representation of the overall topic area, logic frameworks help us to consider interventions that might be used to address important public health issues, set priorities among the interventions, and help us determine which outcomes need to be considered, including important benefits and harms.

The next step is to develop an *analytic framework* (Figure 10–2 is one example of an analytic framework, related to the logic framework in Figure 10–1) for each intervention to be reviewed, which expands on a portion of the logic framework. Analytic frameworks depict in more detail the complex relationships between preventive interventions and outcomes. They are used to map the plan for evaluating interventions, and they guide the search for

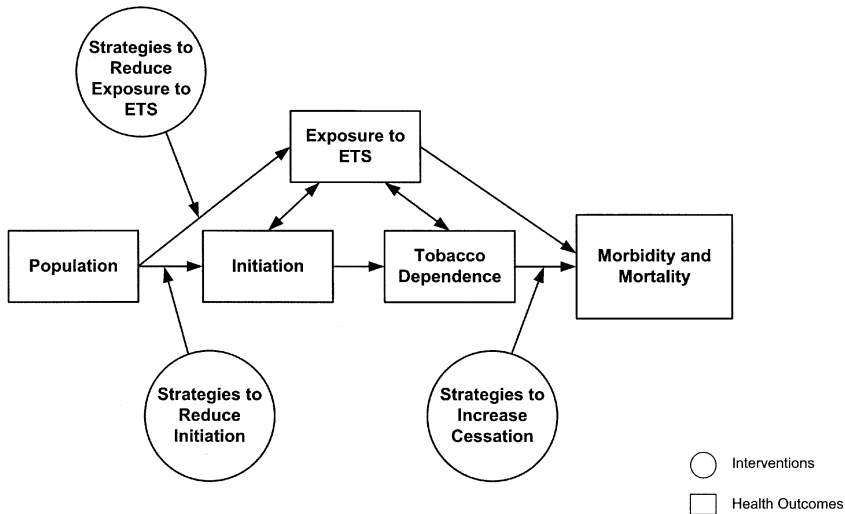


Figure 10–1. Example of a logic framework illustrating the conceptual approach used in the tobacco reviews. (ETS = environmental tobacco smoke.) (Reprinted from *Am J Prev Med*, Vol. 20, No. 2S, Hopkins DP et al., *Reviews of evidence regarding interventions to reduce tobacco use and exposure to environmental tobacco smoke*, p. 18, Copyright 2001, with permission from American Journal of Preventive Medicine.)

evidence. Analytic frameworks are also useful for identifying exactly which interventions will be assessed in the systematic review and which outcomes will be used to assess the success (or effectiveness) of an intervention. Generally, to consider an intervention effective, we require that it either shows improvements in health or leads to changes in behaviors or other factors that have been shown to result in better health. For example, we determined that tobacco use cessation and improved vaccine coverage are clearly related to better health and used them as acceptable markers for success in our reviews. In contrast, knowledge and attitudes are usually considered to be important but are not considered, independently, to translate into improvements in population health. Analytic frameworks are also useful to make certain that we do not overlook important outcomes and that we have targeted the appropriate types of evidence for search and retrieval.

We continue to use the analytic frameworks after a review is completed to communicate our findings. For example, we can use the frameworks to illustrate the numbers of studies and strength of evidence linking interventions to particular outcomes, to highlight gaps in knowledge where further research is needed, and to help users choose interventions that are good matches for particular local problems.

In general, we develop logic and analytic frameworks and intervention categories early in the review process and use them to guide the reviews. However, this process allows for adding important questions as they arise.

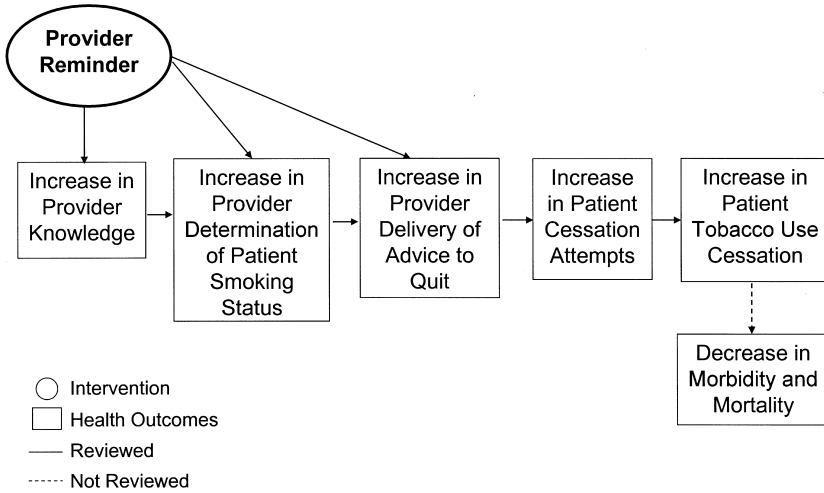


Figure 10–2. Example of an analytic framework illustrating the relationships between the intervention and outcomes. Provider reminders are one of the interventions in the larger category of strategies to increase tobacco use cessation.

Define and Group Interventions

We usually consider four aspects of each intervention:²

1. The type of intervention (including content, activities, and breadth of focus);
2. How the intervention is delivered (e.g., who delivers it; time period, frequency, and duration);
3. Who is targeted (e.g., general population, high-risk group, professional group); and
4. Where the intervention is delivered (e.g., setting).

We also most typically look at the effectiveness of interventions in achieving particular outcomes (e.g., home visiting interventions to promote vaccination probably have different effects from home visiting interventions to prevent violence). The title used for a particular type of intervention usually highlights one or more of these dimensions (e.g., the name “School-Based Education to Reduce Tobacco Use” is based on the setting [schools] and the outcome [tobacco use], although the other dimensions also apply).

To perform *Community Guide* reviews and make recommendations, we define a set of *like* interventions on which to focus a particular review. We determine whether related interventions can be grouped together on the basis of their similarity (in terms of the four dimensions described above and the outcomes addressed) and on whether a single conclusion can be drawn about how the intervention affects one or more outcomes of interest. We

sometimes consider other factors as well, such as depth of available literature or the theoretical basis for making the case that the interventions are substantially similar. The process of grouping interventions can be a challenge because no two community interventions are exactly alike and because nomenclature used in available studies can vary widely.

The strategy for grouping interventions should allow fair, understandable, and valid evaluations of various kinds of interventions. It is important to avoid combining too broadly—that is, comparing results from interventions that differ in important ways in terms of their conceptual basis or the evidence they provide. At the same time, it is important to avoid groupings that are too narrow, which can inhibit our ability to draw conclusions and compare and contrast results. We find that some prudent combination of information about similar but not identical interventions is needed to be able to draw useful conclusions.

Select Interventions to Evaluate

Topics addressed in *Community Guide* systematic reviews are very broad, and the intervention reviews are time-consuming and resource intensive. For these reasons, it is necessary to set priorities for which interventions to review.

The process of selecting interventions for review involves developing the logic framework, deciding whether any areas of the logic framework will be excluded from further consideration (e.g., because an area was already covered in another review), developing a candidate list of interventions, and setting priorities using some type of voting procedure among the team and the consultants. The Task Force approves or modifies the resulting priorities.

Priority-setting criteria are next adapted for a particular review. Teams consider the following issues: potential to reduce the burden of disease and injury; potential to increase healthy behaviors and reduce unhealthy behaviors; potential to increase the implementation of effective but not widely used interventions; potential to phase out widely used, less effective interventions in favor of more effective or more cost-effective options; and current level of interest among providers and decision makers. Other priority-setting criteria may be added as relevant and appropriate.

Occasionally, review teams may engage in formal scoring and weighting of the criteria. One or more rounds of this process results in a prioritized list of interventions. Although this process for selecting interventions is systematic, it is dependent on judgment. A different group of participants might choose a different set of interventions. Nonetheless, selecting topics and interventions using a process that is as broad, open, and transparent as possible should maximize the likelihood that the reviews and the resulting recommendations will be useful to a broad audience.

Systematically Search for and Retrieve Evidence

Once intervention priorities have been set, we engage in a thorough search for relevant information in the scientific literature. Performing this search systematically increases the validity and generalizability of results. Searches are designed to be as comprehensive as possible to reduce the likelihood that the review will evaluate a biased selection of studies.

To search for evidence, we:

- Determine which types of documents are relevant to the study question
- Identify relevant systematic or narrative reviews, and then identify additional studies from the reference sections of these reviews
- Determine which databases are most likely to yield the appropriate document types
- Determine the search parameters and inclusion criteria
- Conduct the search
- Read titles and abstracts of the resulting document list to determine potential relevance
- Obtain potentially relevant documents
- Review documents to confirm that they meet inclusion criteria
- Review documents for additional references

At a minimum, *Community Guide* reviews include recent intervention studies published in journals. Government and other technical reports can also be included. If, after the first attempts at a search, the evidence found for a particular intervention is sparse or nonexistent, other document types, such as dissertations, books, or abstracts, may be considered. This process systematically identifies readily available evidence. It is not possible for any review to include all evidence that might be retrieved if unlimited resources were available.

We are charged with producing public health recommendations applicable in the United States. Including studies from more countries can increase the amount of available literature and may also increase the number of approaches that have been evaluated, but can potentially reduce applicability to the United States. To date, all of our reviews have included U.S. studies, most have included studies from other developed countries, and a few have included studies from developing countries. For most reviews, we have been willing to assume that public health interventions in developed countries might translate to the United States if adapted appropriately. When we have limited ourselves to U.S. studies, we have generally done so because the local social context is believed to have important effects on intervention effectiveness. For example, for some of our reviews of interventions related to violence prevention, the social context was thought to differ substantially between the United States and other developed countries, so we reviewed only

U.S. studies. On the other hand, for most interventions we have argued that the developing world context has been sufficiently different from that of the United States to preclude inclusion of studies from the developing world in our reviews.

Given resource constraints, to date we have limited our reviews to studies published in English. If, in the future, our subject matter experts suggest that additional key references are published in a different language, we will seek to have the relevant studies translated.

Assess Individual Studies

To ensure consistency, reduce bias, and improve validity and reliability in *Community Guide* systematic reviews, a standardized abstraction form and procedure has been developed² and is available at www.thecommunityguide.org/methods/abstractionform.pdf. The form guides reviewers through the process of summarizing studies. It captures important descriptive information about each intervention; identifies qualitative and quantitative results; assesses suitability of study design; and documents important threats to validity.² Each included study is read and summarized by at least two reviewers. When reviewers have differences in opinion, team members reconcile those differences. The abstraction form balances flexibility for the evaluation of papers reflecting different study designs and intervention types with the necessity of a common, systematic approach to the reviews to maximize validity and reliability.

Assess and Summarize Bodies of Evidence

Summarizing Evidence of Effectiveness

We characterize study designs in terms of suitability and quality of execution. Three possible categories describe suitability: greatest, moderate, or least suitable design. These categories are based on the presence or absence of study design characteristics that affect our confidence that the intervention being evaluated really caused the effects or outcomes being measured (internal validity).¹

Quality of study execution is assessed on the basis of six characteristics (descriptions of the study population and the intervention, sampling of the study population, measurement of exposures and outcomes, data analysis, interpretation of results, and other threats that have not already been addressed in the other categories) and limitations in any of these characteristics, as defined by the *Community Guide*,² are noted. (Details of scoring quality limita-

tions are available at www.thecommunityguide.org/methods/scoring-rules.htm.) Each study is categorized as having good, fair, or limited quality of execution based on the number of limitations. Performance in the quality domains can either affect our confidence that the intervention being evaluated really caused the effects or outcomes being measured (internal validity) or our confidence that the study results can be generalized to populations and contexts beyond the particular ones included in the studies themselves (external validity).

To give decision makers information about what to expect, we summarize the information across studies in a variety of ways. Results across a group of related studies are always discussed qualitatively (that is, using tables and text). Sometimes, we perform no other summary. Results are also summarized using graphs and quantitative methods when possible and when such an analysis is useful. In addition, we strive to make a statement about the typical size of the effect and how much that effect has varied across studies. We usually calculate an average effect using an easily understandable effect measure (either the mean or the median) as our best estimate of changes that will occur if the intervention is used.

Determining whether studies can be combined is a judgment that depends on several characteristics of the body of evidence. Sometimes we conclude that *no combination* across studies is feasible or desirable. We would reach this conclusion in cases where there are too few studies; the interventions or components of the interventions are too varied to allow meaningful overarching conclusions; measured outcomes within the body of evidence are too diverse to allow assessment of the overall effectiveness of selected interventions; or the quality of individual studies is too poor or the problems with the body of evidence overall are too severe to allow confidence in any measured effect. It is hoped that this conclusion and its rationale will lead to improvements in future research.

Sometimes, *qualitative combination* is feasible and useful but quantitative summary is not. In this case we cannot produce a mean or median effect size. Such a conclusion might read, for example, “All of the qualifying studies showed improvements in the health outcomes evaluated; however, the diversity of health outcomes studied precluded the calculation of a ‘typical’ effect.” At other times, quantitative summaries are not thought to provide substantial added value for decision making. For example, “All qualifying studies showed increased measures of knowledge” is frequently as useful as “Knowledge measures generally increased by an average of 1 standard deviation with a 95% confidence interval of .6 to 1.6.”

We often use “*simple*” *quantitative combination*. Where quantitative results are available from studies with reasonable quality and conceptual similarity, we find it useful to transform study results to a common scale (e.g.,

percentage point changes). This allows us to describe the size of the changes associated with the interventions and also to describe the distribution of results using plots and simple descriptive statistics (e.g., median and range). This has the added benefit of providing additional empirical data on the similarity or difference in results across studies, thus providing additional data to inform decisions on whether the interventions and the results are sufficiently similar to combine and, if so, to present a *typical* effect size. If the results are not sufficiently similar, we attempt to explain the variations using stratified analyses. For example, studies could be stratified by characteristics of the study (e.g., the length of follow-up), characteristics of the intervention (e.g., duration), characteristics of the population (e.g., sex of the study population), or characteristics of the context (e.g., type of setting in which the intervention occurred).

We sometimes combine study results using the formal statistical procedures of *meta-analysis*, which involves converting study results to a common metric, calculating a confidence interval for each result, and calculating a summary estimate of effect across the group of studies that is weighted by the precision of each study. For example, a study with a narrow confidence interval (higher precision) would typically carry greater weight in the overall average. Additional variables that may affect the study results can be controlled through procedures called *meta-regression*, in which intervention and other study characteristics are used in a statistical model to explain variability in outcomes. The statistical details of these techniques are beyond the scope of this chapter but are available in other sources.⁴

The choice between *simple* quantitative combination and formal statistical procedures involves a range of pros and cons. Simple quantitative combination has the potential advantages of being easier to implement and understand, and of allowing more studies to be included (because we need not exclude studies, for example, that do not report information on the variability of results). Formal statistical procedures also have potential advantages: they may give greater weight to more precise results, and they use formal rules for deciding when an effect is significantly different from no effect. For ease of interpretation and simplicity of calculation, we have tended to err toward use of the simplest analytic strategy that will adequately represent the available data.

Finally, the body of evidence of the effectiveness of an intervention is characterized as strong, sufficient, or insufficient on the basis of the number of available studies, the strength of their design and execution, and the size and consistency of reported effects (Table 10–1).

One can achieve sufficient or strong evidence in a variety of ways. For example, sufficient or strong evidence can be achieved through one or two very well designed and executed studies with few threats to validity. Alternatively,

Table 10–1. Assessing the Strength of a Body of Evidence on Effectiveness of Population-Based Interventions in the *Guide to Community Preventive Services*

| <i>Evidence of Effectiveness^a</i> | <i>Execution (Good or Fair^b)</i> | <i>Design Suitability (Greatest, Moderate, or Least)</i> | <i>Number of Studies</i> | <i>Consistent^c</i> | <i>Effect Size^d</i> | <i>Expert Opinion^e</i> |
|--|--|--|--------------------------|-------------------------------|--------------------------------|-----------------------------------|
| <i>Strong</i> | Good | Greatest | At least 2 | Yes | Sufficient | Not used |
| | Good | Greatest or moderate | At least 5 | Yes | Sufficient | Not used |
| | Good or fair | Greatest | At least 5 | Yes | Sufficient | Not used |
| | (Meets design, execution, number, and consistency criteria for sufficient but not strong evidence) | | | | Large | Not used |
| <i>Sufficient</i> | Good | Greatest | 1 | Not applicable | Sufficient | Not used |
| | Good or fair | Greatest or moderate | At least 3 | Yes | Sufficient | Not used |
| | Good or fair | Greatest, moderate, or least | At least 5 | Yes | Sufficient | Not used |
| <i>Expert Opinion</i> | Varies | Varies | Varies | Varies | Sufficient | Supports a recommendation |
| <i>Insufficient^f</i> | A. Insufficient designs or execution | | B. Too few studies | C. Inconsistent | D. Small | E. Not used |

^aThe categories are not mutually exclusive; a body of evidence meeting criteria for more than one of these should be placed in the highest possible category.

^bStudies with limited execution are not used to assess effectiveness.

^cGenerally consistent in direction and size of effect

^dSufficient and large effect sizes are defined on a case-by-case basis and are based on Task Force opinion.

^eExpert opinion will not be routinely used in the *Community Guide* but can affect the classification of a body of evidence as shown.

^fReasons for a determination that evidence is insufficient will be described as follows: A. Insufficient designs or executions, B. Too few studies, C. Inconsistent, D. Effect size too small, E. Expert opinion not used. These categories are not mutually exclusive, and one or more of them will occur when a body of evidence fails to meet the criteria for strong or sufficient evidence.

Reprinted from *Am J Prev Med*, Vol. 18, No. 1S, Briss PA et al., Developing an evidence-based *Guide to Community Preventive Services*—methods, p. 40, Copyright 2000, with permission from American Journal of Preventive Medicine.

and more commonly, a group of individually less persuasive studies can provide sufficient or strong evidence taken together, especially if their flaws are not overlapping.

Summarizing Other Effects

Interventions designed to lead to health outcomes sometimes result in what we refer to as *other effects*, that is, important outcomes of the intervention that are side effects rather than the primary effects used to assess effectiveness. Other effects may be intentional or incidental and can relate to either health or non-health outcomes. They can include *harms* (for example, sobriety checkpoints may compromise motorist privacy) or *benefits* (for example, workplace smoking bans may reduce the risk of fire and workplace cleaning costs).

Community Guide reviewers identify potentially important other effects and systematically search for and evaluate the strength of evidence supporting these, following the same process used for assessing effectiveness. Identifying important other effects may affect the Task Force recommendations. For example, credible evidence that harms outweigh benefits will lead to recommendations that interventions not be used.

In reviews to date, available data on either the harms or side benefits of community interventions has been sparse. Our efforts to summarize other effects have thus been used mainly in developing directions for future research rather than recommendations.

Assessing the Applicability of Available Effectiveness Data

To help users determine the likelihood that reviewed interventions will apply to their local populations and settings, *Community Guide* reviews provide information on the applicability of bodies of evidence and resulting recommendations. By studying the conceptual basis of the intervention and the variability or robustness of empirical findings across different contexts, we find that we can generally describe applicability using one of these four conclusions.

- Review findings are likely to be applicable across a broad range of settings, populations, or intervention characteristics.
- Review findings are likely to be applicable across a broad range of settings, populations, or intervention characteristics, assuming that the intervention is appropriately adapted to the population of interest.
- Review findings are applicable only to populations or settings that have been included in the studies, and broader applicability is uncertain.

- Review findings are applicable only to the populations or settings included in the studies.

These categories have been developed to provide a structure for decisions and to link the decisions to supporting conceptual and empirical information. They should not be thought of as a hierarchical grading scheme—that is, the first category is not “better” than the last. Rather, they are intended to help users assess the extent to which the findings may apply to their local situations.

Economic Evaluations

Interventions to improve health are typically constrained by scarce or limited resources. Decision makers seek useful information about the resources required for various interventions, and the return that can be expected relative to the cost of an intervention, to help them allocate resources to produce maximum improvements in health. Whenever economic data on recommended interventions are available, the review team collects, abstracts, adjusts, and summarizes results from economic studies to support decision making. We do this to make economic information more available and comparable across studies. (The rationale, utility, procedures, and instruments for summarizing economic information are discussed further in Chapter 11.) However, economic information on public health interventions reviewed to date has been sparse.

Summarizing Barriers to Implementation of Interventions

Community Guide reviews provide information on barriers that might impede implementation of interventions. Examples of such barriers include political opposition to smoking restrictions by smokers and the tobacco industry, difficulty passing legislation on vaccination requirements, or state constitutional prohibitions against sobriety checkpoints to reduce alcohol-impaired driving.

Each recommended intervention evaluated in a *Community Guide* review includes a description of barriers to implementation that have been encountered or that the systematic review development team thinks likely to be encountered when implementing particular interventions. This information can come from reviewed studies, additional evidence searches specific to barriers (although research descriptions of barriers have been sparse), or the opinion of review team members who are subject matter experts. Knowledge of barriers can help decision makers select interventions or help practitioners anticipate potential problems so that they can find ways to work around barriers early in the implementation process.

Translating Evidence into Recommendations

The *Community Guide* provides systematic, evidence-based recommendations to diverse public health audiences. The recommendations highlight which interventions are most appropriate and effective for promoting health and preventing disease and injury in communities based on the aggregated evidence of public health research.

After reviewing the findings from a systematic review, the Task Force either recommends use of the intervention or finds that evidence is insufficient to determine whether or not the intervention is effective. The Task Force's recommendations primarily address evidence of the effectiveness of interventions, but other factors, such as applicability, barriers, and economic evidence, are sometimes incorporated.

Using Effectiveness Data to Formulate Recommendations

Recommendations are primarily driven by effectiveness data, which the *Community Guide* systematically summarizes using the guidelines shown in Table 10–1. Evidence for the effectiveness of interventions is determined to be strong, sufficient, or insufficient to determine effectiveness based on the number of available studies, the suitability of their designs and quality of execution, and the consistency and size of reported effects.

In general, a direct relationship exists between strength of evidence and strength of recommendation, as shown in Table 10–2. Although the recommendation language described here has evolved slightly from previous versions,¹ neither the process of developing recommendations nor the intent of the recommendations has changed.

Table 10–2. Relationship Between Strength of Evidence of Effectiveness and Recommendations

| <i>Strength of Evidence of Effectiveness</i> | <i>Recommendation</i> |
|--|---|
| Strong | The intervention is recommended on the basis of strong evidence of effectiveness |
| Sufficient | The intervention is recommended on the basis of sufficient evidence of effectiveness |
| Insufficient information | Available studies do not provide sufficient evidence to determine the effectiveness of the intervention |
| Sufficient or strong evidence of ineffectiveness or harm | Use of the intervention is discouraged based on sufficient or strong evidence |
| Insufficient empirical information, supplemented by expert opinion | The intervention is recommended on the basis of expert opinion |

The consistency of results also affects recommendations. When evidence of effectiveness is inconsistent and the inconsistency can be attributed to certain characteristics of the population, setting, or intervention, recommendations can be targeted to a specific context. For example, some interventions may be appropriate for urban populations but not for rural populations, or for use in health department clinics but not in managed care organizations.

All else being equal (e.g., strength of evidence and consistency of findings), a large effect size can strengthen a body of evidence. Conversely, a small effect size can weaken a body of evidence. The Task Force also has the option of making a recommendation based solely on expert opinion, but has not done so to date.

Other Factors

As described above, *Community Guide* reviews also collect and evaluate information on the applicability of findings to various populations and settings and on the intervention's other effects (that is, side effects). These factors can affect the recommendations. In contrast, economic impact and barriers that have been observed when implementing interventions, although included in systematic reviews for the consideration of users, do not generally influence Task Force recommendations.

Applicability

The Task Force makes judgments about the contexts in which the recommendations apply. These judgments are based on (1) a conceptual understanding of the intervention in question, (2) what the research evidence says about the consistency or variability of results of the intervention across different intervention characteristics, and (3) the characteristics of the settings and populations. In most cases, *Community Guide* recommendations have been thought to apply to a wide range of populations and settings, but, at times, recommendations are targeted to a specific context. For example, some diabetes interventions have been recommended for people with type 1 but not type 2 diabetes, or vice versa, and standing orders to promote vaccinations have been recommended for adults but not for children.

Other Effects

Documented harms to health that outweigh benefits will lead to recommendations against use of interventions. However, because harms are frequently understudied relative to benefits, postulated serious harms that have not yet been adequately studied may lead to recommendations for further research rather than to practice recommendations (even if the intervention has been found to be effective in changing some outcomes). There may also be

cases in which an intervention is effective in some populations but harmful to one or more other populations. In such cases, the Task Force may make a more narrowly targeted recommendation than would otherwise be made or may recommend against use of the intervention.

Economic Analyses

Economic information does not routinely influence *Community Guide* recommendations, because the availability and quality of data are often limited. Additionally, different users will bring different values to bear in terms of how and whether economic information should be incorporated into decision making.

Barriers to Intervention Implementation

Each intervention evaluated in a *Community Guide* systematic review includes information on barriers that have been encountered when implementing interventions. This information is primarily provided for decision makers to consider when selecting interventions and does not typically influence recommendations.

The Role of Task Force Opinion and Judgment

Although *Community Guide* reviews are explicit and systematic, an element of judgment is always involved. The same is true for the development of recommendations. Many Task Force decisions (including, for example, intervention definitions and outcomes used to define success) influence the resulting recommendations. A different group of decision makers could reach different conclusions.

Another area in which Task Force judgment may be influential is where evidence is insufficient to determine the effectiveness of an intervention. In such cases, the Task Force has reserved the option of making recommendations based solely on expert opinion. However, after more than 170 reviews, the Task Force has not yet elected to exercise this option.

Summarizing Evidence Gaps

Community Guide systematic reviews identify and assess existing evidence to provide a basis for public health decisions. An important additional benefit of these reviews is identification of areas where this evidence is lacking or is of poor quality. Identifying these evidence gaps or research needs— questions that remain to be answered about a given topic area— can help researchers and funding agencies focus their efforts on areas most in need of further study (i.e., research agendas). For each intervention evaluated, whether or not

evidence was already sufficient for a recommendation, we identify remaining evidence gaps. Where evidence of effectiveness of an intervention is sufficient or strong, remaining questions about effectiveness, applicability, other effects, economic consequences, and barriers are summarized. Where evidence of effectiveness of an intervention is insufficient, remaining questions about only effectiveness and other effects are summarized. Applicability issues are summarized only if they affect the assessment of effectiveness. The team decides if it is premature to identify research gaps in economic evaluations or barriers before effectiveness is demonstrated.

For each category of evidence, issues that emerge from the review are identified, based on the informed judgment of the team. Several factors influence that judgment. When a conclusion is drawn about evidence, the team decides if additional issues remain. Specifically:

- If effectiveness was demonstrated by using some but not all outcomes, all other possible outcomes are not necessarily listed as research gaps.
- If the available evidence was thought to be generalizable, we do not necessarily identify all subpopulations or settings where studies had not been done.
- Within each body of evidence, the team considers whether there are general methods issues that would improve future studies in that area.

(See Chapter 12 for a more complete discussion of research needs.)

CONCLUSION

The rigorous systematic reviews and Task Force recommendations in the *Community Guide* are designed to make a vast range of current scientific evidence accessible and useful to decision makers who are responsible for selecting appropriate health interventions for their communities. Information in these reviews and recommendations includes systematically derived and communicated information on:

- High-priority health topics
- Conceptual models that identify important public health outcomes, high-priority interventions to achieve them, and the proposed ways in which the interventions should work
- The empirical studies that have measured the success of these programs, including
 - The quality of those studies, individually and collectively
 - The size and variability of reported effects, including changes in intermediate and health outcomes and costs
 - Likely applicability of the findings across different intervention characteristics and community contexts

- Barriers to implementation, and
- Task Force recommendations about using or not using an intervention.

In this way, the *Community Guide* contributes to the scientific basis for effective public health practice that should be used in conjunction with information about local needs, resources, priorities, and barriers. *Community Guide* recommendations and reviews also help researchers and those who support research to identify gaps and redundancies in what we know and to develop an appropriate public health research agenda.

Acknowledgments

This chapter was written by Peter A. Briss, MD, MPH, Community Guide Branch, Division of Prevention Research and Analytic Methods (DPRAM), Epidemiology Program Office (EPO), Centers for Disease Control and Prevention (CDC), Atlanta, Georgia; Patricia Dolan Mullen, DrPH, Center for Health Promotion and Prevention Research, University of Texas–Houston, School of Public Health; and David P. Hopkins, MD, MPH, Community Guide Branch, DPRAM/EPO/CDC, Atlanta.

References

1. Briss PA, Zaza S, Pappaioanou M, et al. Developing an evidence-based Guide to Community Preventive Services— methods. *Am J Prev Med* 2000;18(1S):35–43.
2. Zaza S, Wright-de Aguero L, Briss PA, et al. Data collection instrument and procedure for systematic reviews in the Guide to Community Preventive Services. *Am J Prev Med* 2000;18(1S):44–74.
3. Zaza S, Lawrence RS, Mahan CS, et al. Scope and organization of the Guide to Community Preventive Services. *Am J Prev Med* 2000;18(1S):27–34.
4. Cooper H, Hedges LV, eds. *The handbook of research synthesis*. New York: Russell Sage Foundation, 1994.