

https://www.cdc.gov/violenceprevention/pdf/understanding_evidence-a.pdf

Understanding Evidence

Part 1: Best Available Research Evidence



*A Guide to the Continuum of
Evidence of Effectiveness*

Understanding Evidence Part 1: Best Available Research Evidence. A Guide to the Continuum of Evidence of Effectiveness is a publication of the National Center for Injury Prevention and Control, Centers for Disease Control and Prevention.

Centers for Disease Control and Prevention

Thomas R. Frieden, MD, MPH
Director

National Center for Injury Prevention and Control

Linda C. Degutis, DrPH, MSN
Director

Division of Violence Prevention

James A. Mercy, PhD
Acting Director

Authors:

Richard W. Puddy, PhD, MPH
Natalie Wilkins, PhD

Suggested citation: Puddy, R. W. & Wilkins, N. (2011). *Understanding Evidence Part 1: Best Available Research Evidence. A Guide to the Continuum of Evidence of Effectiveness*. Atlanta, GA: Centers for Disease Control and Prevention.

Introduction

Increasing emphasis has been placed on the importance of evidence-informed prevention strategies and evidence-based decision making.¹ Definitions of what constitutes “evidence” have been debated,^{2,3,4,5} but most agree that evidence is extremely important for researchers, practitioners, and policy makers charged with the task of making decisions around the funding and implementation of violence prevention strategies.^{5,6}



What is the Purpose of this Guide?

In this guidance document, we aim to explain the purpose and meaning of the *Continuum of Evidence of Effectiveness*, a tool that was developed to facilitate a common understanding of what the Best Available Research Evidence means in the field of violence prevention. This *Continuum* also serves to provide common language for researchers, practitioners, and policy-makers in discussing evidence-based decision making.

Best Available Research Evidence

The Best Available Research Evidence enables researchers, practitioners, and policy-makers to determine whether or not a prevention program, practice, or policy is actually achieving the outcomes it aims to and in the way it intends. The more rigorous a study’s research design, (e.g., randomized control trials, quasi-experimental designs), the more compelling the research evidence.

The Best Available Research Evidence is widely accepted as the most commonly used type of evidence in fields ranging from medicine to psychology. Although increasingly, other forms of evidence related to clinical/practitioner experience/expertise and setting/contextual factors have been recognized as being crucial to the success of prevention efforts for many behavioral health problems, including violence.^{1,2,6,7}

Three types of evidence

Understanding Evidence = controlled trials??

- 1 While the Best Available Research Evidence is important and the focus of this document, it is not the only standard of evidence that is essential in violence prevention work.

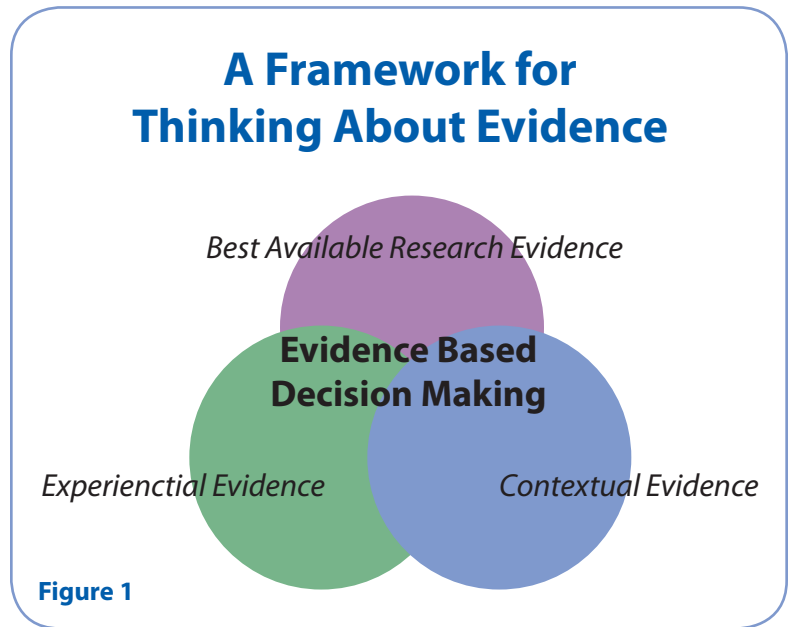
Literature suggests that two other forms of evidence are also very important when making decisions based on evidence:

- 2 • **Experiential Evidence:** This type of evidence is based on the professional insight, understanding, skill, and expertise that is accumulated over time and is often referred to as intuitive or tacit knowledge.⁸
- 3 • **Contextual Evidence:** This type of evidence is based on factors that address whether a strategy is useful, feasible to implement, and accepted by a particular community.^{2,3,5,9}

These three facets of evidence, while distinct, also overlap and are important and necessary aspects of making evidence-based decisions.

As shown in Figure 1, evidence-based decision making occurs when the best available research evidence is combined with the experiential evidence of field-based expertise and contextual evidence.^{1,2,6}

This guide focuses on understanding standards of rigor for the Best Available Research Evidence on violence prevention strategies. The other two facets of evidence, experiential evidence of field-based expertise and contextual considerations, are beyond the scope of the *Continuum* and this guidance document.



Understanding the Continuum

The *Continuum of Evidence of Effectiveness* is a tool that clarifies and defines standards of the Best Available Research Evidence. While in this document, the *Continuum* is applied specifically to the field of violence prevention, it can be used to inform evidence-based decision making in a wide range of health-related areas.

Purpose of the Continuum...

- To present a clear and universal set of standards on the Best Available Research Evidence for the field of violence prevention
- To provide information for decision makers in the field of violence prevention on these standards of the Best Available Research Evidence

The Continuum is intended for...

- Researchers
- Practitioners
- Policy-makers
- Other decision makers in the field of violence prevention

Where Does the Continuum Come from?

The *Continuum of Evidence of Effectiveness* was developed through:

- A thorough review of the literature on evidence
- An examination of existing evidence registries and standards for classification of evidence from the disciplines of psychology, epidemiology, human services, policy, medicine, child welfare, violence, juvenile justice, substance abuse, education, etc. More than 42 sources were considered in the development of the *Continuum*, including:
 - National Registry of Evidence-Based Programs and Practices⁵
 - Blueprints for Violence Prevention¹⁰
 - Community-Based Child Abuse Prevention Programming¹¹
 - Kauffman Best Practices Project¹²
 - Handbook of Injury and Violence Prevention¹³
 - Guide to Community Preventive Services¹⁴
 - California Evidence-Based Clearinghouse¹⁵
 - What Works Clearinghouse¹⁶
 - Find Youth Info¹⁷
 - Promising Practices Network for Children, Families, and Communities¹⁸
 - Violence Prevention: the Evidence¹⁹

These sources all play an important role in providing information about the best available research evidence in their respective disciplines. One shortfall of having so many different registries/sources is the lack of consistency between them in the language, structure, scope, and definitions of the best available research evidence.

The *Continuum* aims to synthesize the information from these registries and create a common understanding of the best available research evidence.

- Expert opinions on the *Continuum* were gathered from researchers, practitioners and policy-makers from a variety of violence-related content areas including: youth violence, self-directed violence, intimate partner violence, sexual violence, and child maltreatment.
- The (horizontal) dimensions that are listed along the left side of the *Continuum* were developed by examining all the domains from existing sources and retaining those with the highest frequency of occurrence.
- The (vertical) areas that are listed across the top of the *Continuum* were developed by examining the various rating systems / levels from existing sources and aggregating them to obtain the most expansive list of categories.

What the Continuum IS

- An educational tool that provides information on the Best Available Research Evidence, the first “sphere” in The Framework for Thinking About Evidence (as shown in Figure 1 on page 4).
- A means of clarifying and defining standards of rigor across the key dimensions that make up the Best Available Research Evidence (i.e., effectiveness, internal validity, etc.).

What the Continuum IS NOT

- A tool for classifying violence prevention programs. In fact, while the *Continuum* specifies different areas of evidence (ranging from “Well-Supported” on the left to “Harmful” on the right), it should be noted that each of these areas depict a “typical” or “ideal” set of standards that should be met, and are unlikely (and not meant) to suggest that specific violence prevention programs will meet every standard.
- A “one stop shop” for evidence-informed decision making.

The *Continuum* does not represent experiential evidence from field-based expertise, nor does it comprehensively address the contextual considerations involved in evidence-based decision-making.

Beyond the Scope of the Continuum

While the *Continuum* broadly addresses context by assessing whether or not a program or policy has been implemented and demonstrated preventive effects in a community setting (External and Ecological Validity dimension), it does not fully address contextual considerations such as:

- **Feasibility**—Can it be successful given the resources available and the economic, social, geographic, and historical aspects of the current setting?
- **Acceptability**—Will it be accepted by the people and decision makers in the current setting?
- **Utility**—Is it useful for the needs of the people in the current setting? Is it appropriate?

Other tools and processes, beyond the *Continuum*, should be used to elicit evidence from field-based expertise and address contextual considerations more rigorously in the evidence-based decision-making process. The CDC is preparing tools and guidance documents on both experiential and contextual evidence, which will be available in the future for researchers, practitioners, and policy-makers to use in the evidence-based decision making process. Until these tools are made available, the resources listed on page 7 may be helpful in understanding evidence-based decision making more broadly, including information on experiential and contextual evidence.

Additional Resources

National Collaborating Center for Methods and Tools

<http://www.nccmt.ca>

Key word search: "Evidence-Informed Public Health"

Canadian Health Services Research Foundation

<http://www.chsrf.ca/Home.aspx>

Key word search: "Conceptualizing and combining evidence for health system guidance" and "What counts? Interpreting evidence based decision-making for management and policy"

Annual Review of Public Health

<http://www.annualreviews.org/journal/publhealth>

Key word search: "Evidence-Based Public Health: A Fundamental Concept for Public Health Practice"

How Can the Continuum be Used?

The *Continuum* is designed to be used as a tool to help researchers, practitioners, and policy-makers better understand best available research evidence, and why this evidence is important. On a practical level, the *Continuum* can be used to help practitioners and policy-makers make decisions about which violence prevention strategies to adopt in their communities. It can also be used by researchers and practitioners to identify which aspects of a prevention program, practice, or policy can be improved to better demonstrate the evidence of its effectiveness.

Example: A local policy-maker is charged with the task of creating a teen dating violence prevention policy that would be implemented in middle and high schools throughout his/her district. The *Continuum* can be used by this policy maker to determine the strength of the evidence of effectiveness for a number of different school-based teen dating violence prevention policies under the most rigorous standards of evidence.

Example: The director of a community-based after-school program wants to show students, parents, and community leaders the impact a program is having in preventing youth violence. The *Continuum* can be used by this prevention practitioner to determine which aspects of the program and its evaluation need to be improved to better demonstrate its impact.

Example: A researcher is asked to evaluate a University's suicide prevention initiative. The *Continuum* can be used by this researcher to explain ways that the University can make the initiative more rigorous in its design, so that more compelling evidence of its effectiveness may be gathered.

Strength of Evidence and Effectiveness

The *Continuum* is based on two underlying facets of the Best Available Research Evidence:

1. Strength of Evidence

- How rigorously has a program, practice, or policy been evaluated?
- How strong is the evidence in determining that the program or policy is producing the desired outcomes?
- How much evidence exists to determine that something *other* than this program or policy is responsible for producing the desired outcomes?

2. Effectiveness

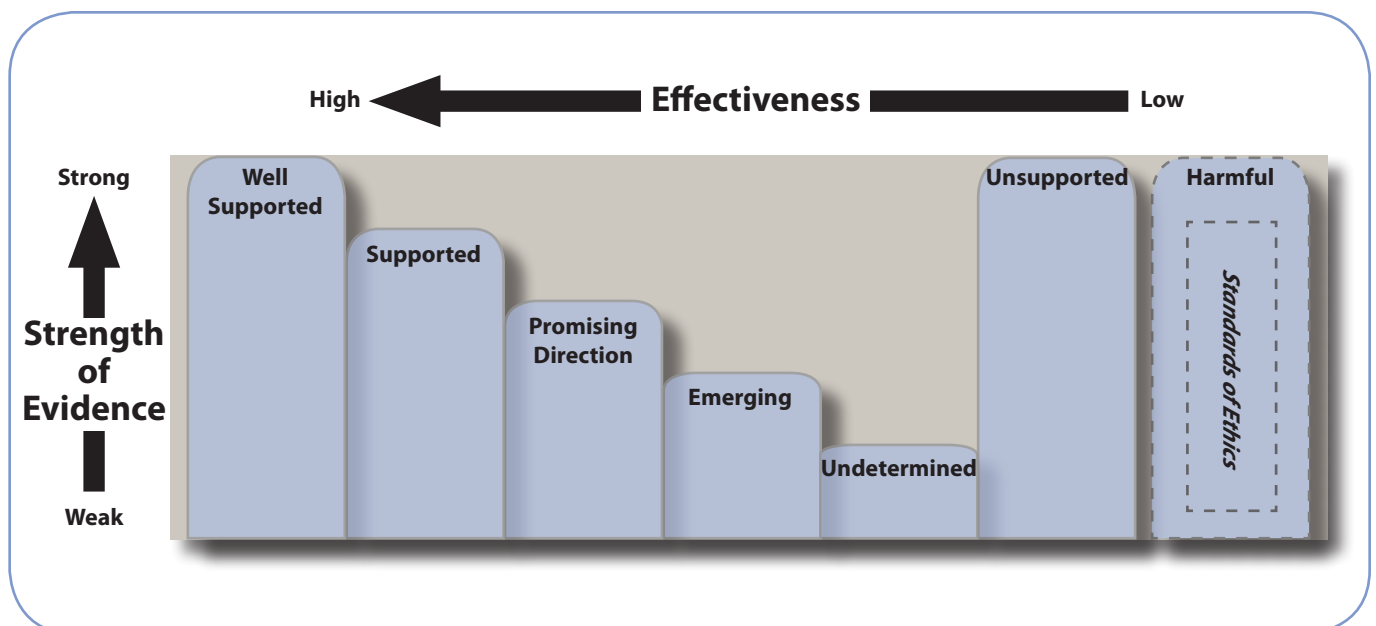
- Is this program, practice, or policy producing desired outcomes?
- Is it producing non-desirable outcomes?

As shown below, the areas of the *Continuum* are differentiated from one another based on the extent to which they meet the requirements of these two facets. These areas range from highly rigorous and effective (“Well-Supported”), to highly rigorous, yet ineffective (“Unsupported”).

NOTE:

The standards used to measure evidence of harm are different from the standards used in other areas of the *Continuum*. In congruence with the ethical standards of health and social sciences,^{20,21} any indication that a program, practice, or policy has a harmful effect on participants, regardless of a study’s rigor (internal validity or research design), is considered strong enough evidence to classify it as harmful.

It is also important to note that the areas of the *Continuum* are meant to represent an “accumulative effect” of the strength of evidence. Starting at the “Undetermined” area and moving left, each area is considered to uphold the standards of evidence described in the areas to its right as well as the additional rigor and standards of evidence specified within its own area on the *Continuum*.



I am having trouble making sense of this continuum!

Continuum of Evidence of Effectiveness

	Well Supported Supported		Promising Direction / Emerging / Undetermined More Research Needed			Unsupported	Harmful	
Effect	Found to be effective		Some evidence of effectiveness	Expected preventive effect	Effect is undetermined	Ineffective	Practice constitutes risk of harm	
Internal validity	True experimental design	Quasi experimental design	Non-experimental design	Sound theory only	No research No sound theory	True or quasi experimental design	Any design with results indicating negative effect	
Type of evidence/ research design	Randomized control trials and meta-analysis / systematic review		Quasi experimental design	Single group design	Exploratory study	Anecdotal / Needs assessment	Randomized control trials or quasi experimental design	Any design with results indicating negative effect
Independent replication	Program replication with evaluation replication		Program replication without evaluation replication	Partial program replication without evaluation replication		Program replication with evaluation replication	Possible program replication with/without evaluation replication	
Implementation guidance	Comprehensive		Partial	None		Comprehensive	Comprehensive/partial	
External and ecological validity	Applied studies—different settings (2+)	Applied studies—similar settings (2+)	Real-world informed	Somewhat real-world informed	Not real-world informed	Applied studies—same/different settings	Possible applied studies—similar/different settings	

NOTE:

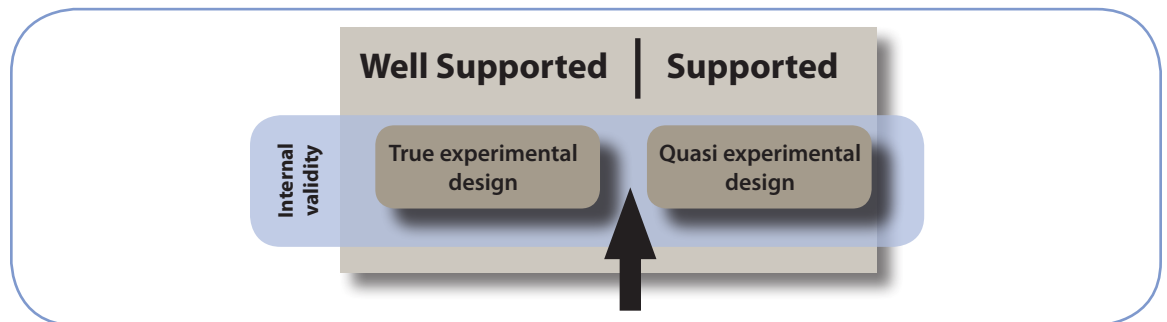
Not all violence prevention strategies will reach the “Well-Supported” level on all dimensions of the *Continuum*.¹³

Example: Randomly assigning participants to a treatment or control group may not be feasible or acceptable in certain settings. In this case, using a quasi-experimental design with a more naturally occurring comparison group (e.g., a wait-list) may be more feasible and acceptable, even though this may mean that the program does not meet the threshold of evidence needed to be considered “Well Supported” under the Internal Validity and Research Design Dimensions of the *Continuum*.

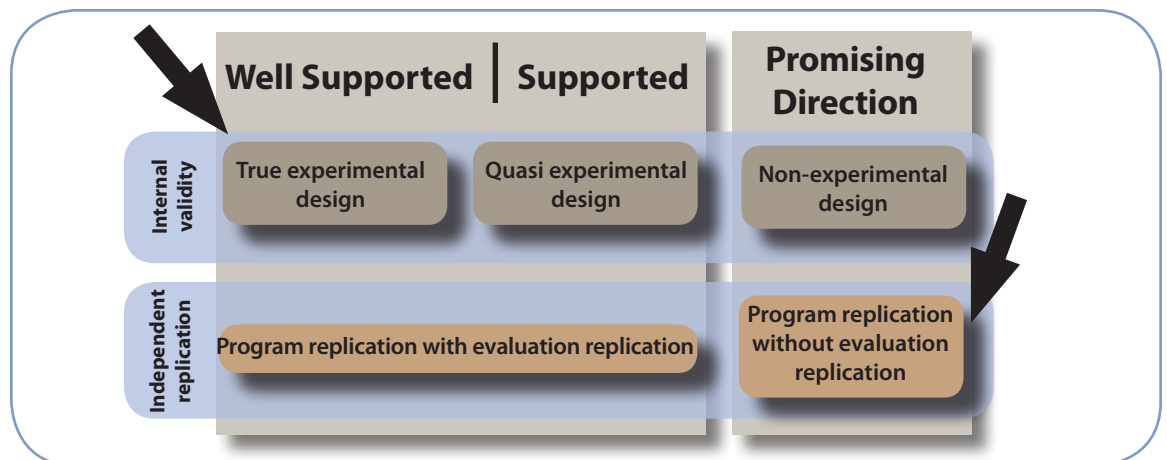
Flexibility of Continuum Areas

The *Continuum* is not meant to be used to classify violence prevention programs, and the areas of the *Continuum* represent “ideal” or “typical” depictions of varying degrees of the best available research evidence. The spaces between these areas on the *Continuum* are meant to show that in reality, prevention programs, practices, and policies may fall between one or more of these areas.

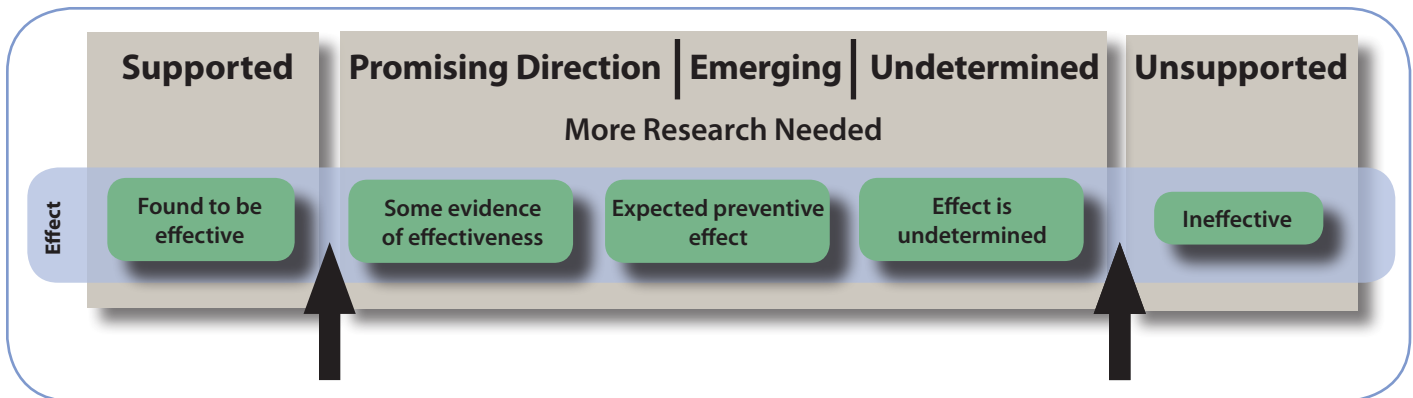
Example 1: A prevention program may have been rigorously evaluated in a randomized control trial (RCT), which is the strongest design in terms of internal validity (relative to other designs). However, even with RCT’s, there still could be other potential threats to internal validity that would place a program between the “Well Supported” and “Supported” areas of the *Continuum*.



Example 2: A prevention practice may fall in one area of the *Continuum* on one dimension, and in another area for another dimension. For example, a practice may be very rigorously evaluated and fall in the “Well Supported” area for the internal validity dimension, but it may not have been independently replicated and thus fall in one of the weaker areas such as “Promising Direction” for the independent replication dimension.



Example 3: The open spaces between the “Supported and Promising Direction,” and “Undetermined and Unsupported” areas of the *Continuum* are meant to separate the areas describing the highest standards or research-based evidence from the areas in which more research is needed to determine effectiveness.



Dimensions of the Continuum

The *Continuum of Evidence of Effectiveness* is made up of six dimensions, each of which addresses a specific aspect of the best available research evidence:

- Effect
- Internal Validity
- Research Design
- Independent Replication
- Implementation Guidance
- External and Ecological Validity

What follows is an explanation of each dimension and why it is important for determining the effectiveness and scientific rigor of prevention strategies.

Effect

The effectiveness of a violence prevention strategy is based on the strategy’s ability to reduce violence-related outcomes. The most effective strategies produce preventive effects in the short term, long term, or both.²²

Short Term Outcomes/Preventive Effects

- Reduce violence-related behaviors (e.g., physical fighting, weapon-carrying, other perpetration behaviors) and injuries
- Increase known protective factors related to violence (e.g., school connectedness)

Long Term Outcomes/Preventive Effects

- Reduce population rates of violence-related injuries, assaults, and homicides
- Reduce rates of school drop-out
- Reduce rates of adult disease and illness (e.g., diseases and illness associated with a history of child maltreatment or intimate partner violence)

Both Short and Long Term Outcomes/Preventive Effects

- Reduce occurrences of violence-related behaviors and injuries and population rates of violence.

Why is Effectiveness important?

Effectiveness is important because it tells us whether a prevention strategy is having an impact on the outcomes of interest. Practitioners must be careful, however, to make sure that the short term outcomes and/or long term outcomes are appropriate for the scope of the strategy and for the violence-related issue it is addressing.

Caveats & Considerations

The effectiveness of a strategy is based on its intent and design. For example, a program designed to modify violence-related behavior would be considered effective if it produced significant outcomes in reducing physical fighting, even if it did not produce significant long term outcomes (e.g., reduction in population rates of assault or homicide). Also, a program may be highly effective in one setting, but not as effective in another setting or context. Practitioners must take into account contextual factors when applying a prevention strategy to a new setting.

Effect on the Continuum...

Found to be Effective

Prevention strategies that are found to be effective are those that are based on sound theory, have been evaluated in at least two, well-conducted studies (this means studies with either a true or quasi-experimental design), and have demonstrated significant, short term and/or long term preventive effects, depending on intent and design. If there are several well-conducted studies all showing the same preventive effects, then the evidence is even more compelling. If there are a number of rigorous evaluations indicating no influence on the short-term and/or long-term outcomes, or any indication that it may cause harm, then it is not considered effective.

Some Evidence of Effectiveness

Some programs may not have two or more rigorous evaluations to demonstrate short and/or long term preventive effects, but they are based on sound theory and have been rigorously evaluated, and the results indicate that they may produce preventive outcomes. These programs show some evidence of effectiveness, although these effects cannot be considered as compelling as those that have been subjected to two or more rigorous evaluations and show short and/or long term preventive effects.

Expected Preventive Effect

Some programs may be grounded in theory and have been evaluated with a less rigorous design, or may have been evaluated for short/long-term preventive effects that are different from the outcomes of interest (e.g., program that has shown preventive effects for substance abuse, but hasn't been evaluated for reducing the perpetration of intimate partner violence). These are indications that the program should have an expected preventive effect.

Effect is Undetermined

Prevention programs that have not been evaluated, or have been evaluated poorly — evaluations with neither a true nor quasi-experimental design — whether or not they are based on sound theory, are considered to have undetermined effectiveness. It is not known whether these programs produce short term and/or long term preventive effects.

Ineffective

Ineffective strategies are those that have been evaluated in at least two, well-conducted studies (this means studies with either a true or quasi-experimental design), and have demonstrated no significant short term or long term outcomes in these evaluation studies. In other words, the findings from these rigorous evaluations show that they do not change or reduce violent behavior.

Practice Constitutes Risk of Harm

A prevention strategy is considered to be harmful if there is an indication that it causes harmful outcomes. This includes short term outcomes, long term outcomes, and/or unexpected outcomes. These harmful outcomes may be due to the inherent nature of the program (i.e., something about the program itself causes harm), its implementation (the way it was delivered/carried out causes harm), an interaction with certain population-related factors (e.g., causes harm for individuals with specific characteristics), or an interaction with certain context/setting-related factors (e.g., causes harm in settings with specific characteristics). Prevention strategies that demonstrate harmful effects should not be replicated.

Internal Validity

Internal validity refers to the extent to which the short term and/or long term outcomes of a program, practice, or policy (as mentioned previously) can truly be attributed to it or if these outcomes could have been caused by something else.²³

Example: A school-based violence prevention program focuses on educating adolescents about pro-social conflict resolution in order to decrease violence. The school also makes major changes to the physical structure of the school (i.e., increased lighting and visibility) in order to decrease incidents of violence. In this case, could reductions in violence-related outcomes really be attributed to the pro-social conflict resolution strategy? Could these changes in violence also be attributed to the physical changes made to the school? To address internal validity, the evaluation of the violence prevention program would have to use a research design (see the Research Design section on page 16 for more information) that enables the researchers/practitioners to determine whether reductions in violence are due to the violence prevention program, physical changes to the school, neither, or both.

Why is Internal Validity important?

Internal validity is important because it enables us to determine whether or not outcomes are really due to the program, practice, or policy itself, or if these outcomes could have been produced by something else.²³ The higher the internal validity, the more confidently we can claim that a program is truly producing the effects.

How do you increase Internal Validity?

There are three main things that increase a program's internal validity:

1. A control or comparison group
2. Multiple measurement points
3. Gathering information on other things that could influence outcomes

Control or Comparison Groups

Control or comparison groups do not receive the program, but are tested on short term outcomes and perhaps long term outcomes, depending on the length of follow-up. Having a control or comparison group enables you to test whether those who receive it demonstrate different outcomes from those who do not receive it. If participants who receive the program have significantly better outcomes than those who do not receive it, then it is possible that the program can be credited with producing these outcomes, and not some other factor (although these other factors still need to be measured- as explained below in the section on "Gathering Information on Other Things That Could Affect Outcomes").

In true experiments, participants are randomly assigned to either the control group or the treatment group (group receiving the program). Random assignment significantly increases internal validity because it attempts to remove any systematic differences between the two groups and makes it probable that both groups were the same from the start. Comparison groups may be used in other types of designs as well (e.g., quasi-experimental). However, without random assignment, there may be some selection bias (i.e., the comparison group might not be comparable to the group receiving a program). In choosing a comparison group it is important to select one that is either matched or very similar to the treatment group. This process of establishing that treatment and comparison groups are equivalent from the start decreases the likelihood that other factors are influencing the outcomes of interest.

Multiple Measurement Points

When outcomes are measured at multiple time points, we are better able to determine the influence of a particular prevention strategy as expected. For example, if participants rate high on a risk behavior before they receive a program, then rate lower on that behavior after they have received it, while the control group remains the same, then it is likely that program can be credited with producing this reduction (assuming there are no other threats to validity, such as test-retest effects. For more information on other threats to internal validity see Hoyle, Harris, & Judd, 2001). Measuring change at multiple points in time also allows you to determine whether the observed changes between the groups are maintained over time (e.g., in the months or years following exposure to the program). Gathering information on program outcomes at multiple time points enables you to determine whether differences between the treatment and control groups are maintained, grow wider, or grow narrower over time.

Gathering Information on Other Things That Could Affect Outcomes

There are many things that can influence prevention outcomes besides the program, practice, or policy of interest. This is demonstrated in the previous example of a school-based violence

prevention program that is implemented in a school that has also made major changes to its physical structure (i.e., increased lighting and visibility) in order to decrease incidents of violence. In this case, reductions in violence-related outcomes may be related to the violence prevention program, but could also be related to the physical changes made to the school.

In order to be sure that the prevention strategy of interest is actually the cause of preventive effects and outcomes, other things (such as the changes in physical structure in the previous example) must be measured and taken into account when interpreting outcomes.

Research designs that include control or comparison groups help to measure these other factors. The table below illustrates a research design that could be used in the previous example to determine whether reductions in violence were due to the violence prevention program, the physical changes to the school, neither, or both. In this example, three schools, which are similar on a variety of school characteristics (e.g., size, composition, physical structure, student/teacher staffing ratios, percentage of students receiving free school lunch, etc.), are randomly assigned to a condition. School A implements a violence prevention program and makes changes to the physical structure. School B makes the same changes to the physical structure as School A, but does not implement the violence prevention program. School C receives no program and makes no physical changes. The table shows that School A (treatment school that receives the program and physical changes) shows significant decreases in incidents of violence while the comparison schools (School B receives only physical changes, and School C receives no program and makes no physical changes) do not. This suggests that preventive effects in School A are likely due to the violence prevention program, and not to the physical changes made to the school.

	Violence Prevention Program	Physical Changes	Decrease in Incidents of Violence
School A	Yes	Yes	Yes
School B	No	Yes	No
School C	No	No	No

Internal Validity on the Continuum...

True Experiment

True experiments are considered to be highest in internal validity because participants are randomly assigned to the treatment and control conditions. This helps assess whether the program, practice, or policy is likely responsible for changes in outcomes or if something else could be causing them. The strongest experimental designs also have multiple measurement points (e.g., longitudinal design). These experiments are able to measure not only differences in outcomes between treatment and control groups, but also changes in outcomes over time. This helps to assess whether the demonstrated effects are sustained over time.

Quasi-Experimental

Quasi-experiments are also considered to have high internal validity, although less so than true experiments. Quasi experiments are based on sound theory and typically have comparison groups (but no random assignment of participants to condition) and/or multiple measurement points (e.g., pre-post measures, longitudinal design).

Some quasi-experimental designs (e.g., interrupted time-series) are used to evaluate policy changes or naturally occurring experiments. These evaluations may not have a comparison group but include multiple waves of observation both before and after the introduction of a treatment (e.g., policy change).

Non-Experimental

Relative to experimental and quasi-experimental designs, non-experimental studies are the weakest of the three in terms of internal validity. Even though these designs are not as rigorous as true and quasi-experiments, they may still be based on sound theory and include some empirical aspects geared toward internal validity. Studies that are non-experimental do not have a control/comparison group or multiple measurement points making it difficult to attribute observed changes to the program. An example of a non-experimental study would be one with a single (treatment) group and a pre-post test or a post test only.

Sound Theory Only

Prevention programs based on sound theory only are also unable to establish or attribute observed changes to the program as those based on experimental or quasi-experimental studies. These programs are often exploratory in nature and are rooted in well established research and subject matter expert opinion, suggesting that the program and/or its components may modify known risk/protective factors and produce preventive outcomes.

No Research, No Sound Theory

Programs, practices, and policies that are not based on research or sound theory are considered to be weakest of all in terms of establishing an empirical link to a preventive outcome. In the absence of research or sound theory, there is no evidence to suggest that they are likely to modify known risk/protective factors or produce preventive outcomes. Some, however, may have face validity. This type of validity is concerned with how a measure or procedure appears and whether it seems reasonably well designed and reliable. Unlike other forms of validity, face validity does not depend on established theories for support.

Research Design

Effectiveness is typically measured in a research study. The nature of the design of the research study determines whether and how well we can answer our research question(s) related to effectiveness. The components or elements of these evaluations (such as measures, selection of participants, assignment to group, assessment of outcomes over time, etc.) are known as the research design. The more rigorous the research design, the higher its internal validity and the more likely outcomes can be attributed to the program, practice, or policy instead of something else.¹⁸

Why is Research Design important?

The type of research design used to evaluate a program is important because it determines how well we are measuring its effectiveness. The more rigorous the research design, the better we can interpret outcomes and the more confident we can be that we are accurately measuring its effectiveness.²⁴ Like internal validity, the more rigorous a study's research design, the better we are able to determine effectiveness of a program and be sure that there is not some other explanation for measured outcomes.

Research Designs on the Continuum...

Randomized Control Trial

Randomized control trials are true experiments and are considered to be a highly rigorous research design. They are the strongest research design for establishing a cause-effect relationship. Randomized control trials have a control (no treatment) group and randomly assign participants to the control or treatment condition. Programs that have been implemented and rigorously evaluated multiple times may be examined further in a systematic review or meta-analysis, which provide even more rigorous information on its effectiveness.

Systematic reviews collect information from a number of scientific studies on a specific topic for the purpose of summarizing, analyzing, and interpreting the overall scientific findings on that topic.¹⁴ A meta-analysis is a type of systematic review that uses statistical analyses to combine and analyze the data from single scientific studies on a specific topic and uses these combined findings to generate a single estimate or effect size to make more conclusive statements about the topic.²⁵ The strongest reviews are conducted independently (by a separate entity), consist of studies that were conducted independent from one another, consist of studies that are comparable (similar samples, methods, procedures), and include some form of empirical analysis to draw broader, general conclusions about the effectiveness of a strategy.

Quasi-Experimental Design

If a design uses multiple groups (without random assignment) or includes multiple measurement points, it is considered quasi-experimental. Quasi-experimental designs are considered to be rigorous designs, although not as rigorous as randomized control trials because participants are not randomly assigned to treatment and control conditions and may not be equivalent from the start. In this respect, they are weaker in controlling threats to internal validity than randomized control trials.

Single Group Design

The single group design is not considered as rigorous as the randomized control trial or quasi-experimental designs because it does not include a control or comparison group. Single group designs may also have just one post-measure or they may include a pre and post measure.

Exploratory Studies

Exploratory studies are focused on learning about a program, practice, or policy and the phenomena it addresses. Exploratory studies are based on sound theory derived from prior research and/or knowledge from subject matter experts. The information gleaned from an exploratory study may point to risk and protective factors that are potentially important

to consider in developing or refining a prevention strategy or its components. Examples of methods used in exploratory studies include ethnography, focus groups, sociometrics, and narrative analysis. Some descriptive and observational studies may also be considered exploratory studies.

Anecdotal/Needs Assessment

Studies not based on empirical research or sound theory are the weakest with respect to research design. Studies that are based on anecdotal information (information not derived from empirical research or subject matter expert opinion), needs assessments, or windshield surveys are examples of this kind of research.

Independent Replication

The independent replication of a program involves implementing it with other participants (e.g., in a different school with other students). This replication should be independent, meaning it should be implemented and evaluated by researchers/practitioners who are unaffiliated with the original program and who do not have any conflicts of interest in implementing or evaluating it.

The purpose of independent replication is to determine whether or not a prevention program can: 1) be implemented with other participants, and 2) produce the same effects. Independent replications are not used to determine whether a program can be successfully generalized to a broad variety of settings or populations, just whether it can be replicated. As such, independent replications of programs typically occur with populations that are similar to the original program.²³

Why is Independent Replication important?

Independent replication is important because it tells us whether a prevention program can be repeated and still be effective. Replication helps to establish the strength of a program and its preventive effects and demonstrates that it can be successfully implemented with other participants.

Independent Replication on the Continuum...

Program Replication with Evaluation Replication

Programs that demonstrate the most reliability (ability to repeatedly produce the preventive effects) are those that have been replicated at least once by independent practitioners/researchers, in a similar setting to the original program, using a rigorous research design (randomized control trial or quasi-experimental design), and with high fidelity to the original program (i.e., conducted in the same way as the original evaluation of the program).

Caveats & Considerations

Programs that have been successfully independently replicated should not necessarily be implemented in all types of settings or with very different populations. Factors such as a program's external and ecological validity (see page 20) and other considerations beyond the scope of this document (such as feasibility, acceptability, and utility) must also be taken into consideration when implementing a prevention strategy in a setting that is very different from the original setting. This ensures that the program is appropriate for this different setting and population.¹

Program Replication without Evaluation Replication

Programs that demonstrate some reliability are those that are implemented with high fidelity to the original program and in settings that are similar to the setting of the original program (e.g., a different school with other students). These replications may or may not be conducted by independent researchers/practitioners. Finally, these replications have not been evaluated in the same way as the original evaluation of the program.

Partial Program Replication without Evaluation Replication

Programs that demonstrate weak reliability are those that are partially replicated and have not been evaluated. These replications may or may not be conducted by independent researchers/practitioners. Programs that are the weakest in reliability are those that are not replicated at all since there is no way to measure their reliability.

Possible Program Replication with/without Evaluation Replication

If a program demonstrates harmful effects, it should not be replicated. In some cases, harmful effects may not have occurred during the original implementation of a prevention strategy but may occur in its replication. Evaluations may or may not have been conducted of this replication since a formal evaluation is not needed to prove harm. Once harmful effects have been associated with a program (either in the original or during a replication) no subsequent replications should be conducted.

Implementation Guidance

Implementation guidance includes any and all services and/or materials that aid in the implementation of a prevention strategy in a different setting, including but not limited to: training, coaching, technical assistance, support materials, organizational/systems change consultation, and manuals/guides.²⁶ Implementation guidance is typically created by the original developers of a program in order to help researchers/practitioners implement it appropriately in their own setting.

Why is Implementation Guidance important?

Implementation guidance is important because programs are not likely to be established and carried out appropriately without guidance on how to do so. If researchers/practitioners do not have guidance on how to implement a program, it is likely that it will not have high fidelity, meaning that it will probably not be carried out in the way it was intended.²⁶ Programs that follow all of the comprehensive implementation guidance are more likely to have high fidelity and therefore outcomes can be attributed more confidently to the program itself and not to implementation factors. On the other hand, if implementation support, services, and materials are not available and/or used, there is a chance that implementation issues (not the program itself) may have led to weak or poor outcomes.⁶

Caveats & Considerations

It is important to note that the presence of implementation guidance does not guarantee that a program will be implemented with high fidelity. Practitioners/researchers may not follow this guidance or the organizations and communities within which the program is being implemented may not have the capacity or support to implement it with high fidelity.

Implementation Guidance on the Continuum...

Comprehensive

Comprehensive guidance is the most effective way of ensuring that a program is carried out with fidelity in a different setting. This entails availability and accessibility of any products, services, or activities that facilitate proper implementation in a new setting. These products and services include training, coaching, technical assistance, support materials, organizational/systems change consultation, and manuals/guides, and may be offered by the program's developers or some other entity.

Partial

For some programs, there may be some products, services, or activities to help researchers/practitioners implement them in different settings, but they may be limited in their availability and accessibility. It is important to note that since implementation support and guidance are limited for these programs, there is a chance that implementation issues (not the program itself) may be influencing outcomes.

None

Programs that do not have any products, services, or activities available to help researchers/practitioners implement them in a different setting run a high risk of experiencing implementation issues. This also means there is a significant chance that implementation issues (not the program itself) may be influencing outcomes.

External and Ecological Validity

This dimension of the *Continuum* combines aspects of the principles of external and ecological validity that are relevant to the best available research evidence. External validity refers to whether a program, practice, or policy can demonstrate preventive effects among a wide range of populations and contexts.²³ For example, a parenting skills training program designed to prevent child maltreatment that demonstrated preventive effects in both urban and rural areas with different populations of parents would have high external validity.

Ecological validity, on the other hand, refers to whether the program components and procedures approximate the "real life" conditions of a specific setting.²³ For example, a school violence prevention program that has shown preventive effects in school settings where real world factors (e.g., staff or classroom curriculum changes) exist. On the *Continuum*, external and ecological validity are defined as the extent to which a program has been implemented in the "real world" and has been shown to work in a variety of different applied settings and populations.

Caveats & Considerations

While multiple applied studies demonstrating preventive effects are an indication that a program may be high in external validity, it is important not to make assumptions about whether or not it should/could be successfully implemented in any setting. Contextual factors (i.e., feasibility, acceptability, utility) must also be taken into account when assessing the appropriateness of a program for a given setting.¹

Why are External and Ecological Validity important?

External and ecological validity are important because they tell us whether or not a program works across different applied (“real world”) settings. By implementing the program in a number of different settings we can begin to learn which aspects seem to work with a wide variety of populations and contexts and which aspects may need to be adapted to fit specific settings and populations.

External & Ecological Validity on the Continuum...

Two or More Applied Studies—Different Settings

Programs that demonstrate the highest external and ecological validity are those that have been implemented in two or more applied (“real world”) settings that are both distinct from the original setting and each other in terms of their populations and physical/geographical locations.

Two or More Applied Studies—Same Settings

Some programs have been implemented in two or more applied (“real world”) settings that are similar to one another with similar populations. These prevention strategies demonstrate moderate external and ecological validity although not as much as those implemented in two or more settings that are different and that have different populations.

Real World-Informed

Programs that have not been implemented in applied settings may still demonstrate some external and ecological validity if they are made up of components that are consistent with an applied setting (i.e., using materials and resources that would be available/appropriate in an applied setting). Likewise, programs may demonstrate external and ecological validity if they are implemented in ways that mirror conditions of the “real world” (i.e., deliver the strategy in ways that it would have to be delivered in the real world).

Somewhat Real World-Informed

Some programs have not been implemented in applied settings and are not structured and implemented in ways that are completely consistent with an applied setting. These prevention strategies demonstrate some external and ecological validity if some of their components and implementation approximate conditions in the “real world.”

Not Real World-Informed

Programs that demonstrate the least amount of external and ecological validity are those whose basic components are not consistent with an applied setting and are not implemented in ways that mirror conditions of the “real world.” While it is not known whether these programs will be effective in applied settings, there is no way to measure which aspects work well across different settings and populations or which aspects are setting-specific.

Possible Applied Studies in Similar/Different Settings

Programs that demonstrate harm in any kind of a setting (applied or otherwise) are considered to be harmful. In other words, the program is considered harmful regardless of whether or not it has been conducted in an applied setting or not.

References

1. American Psychological Association (APA, 2005). *Report of the 2005 Presidential Task Force on Evidence-Based Practice*. Washington, DC: American Psychological Association.
2. Canadian Health Services Research Foundation (March, 2004). *What counts? Interpreting evidence based decision-making for management and policy. Report of the 6th CHSRF Annual Invitational Workshop*. Vancouver, British Columbia: Canadian Health Services Research Foundation.
3. Lomas, J., Culyer, T., McCutcheon, C., McAuley, L., & Law, S. (2005). *Conceptualizing and combining evidence for health system guidance*. Ottawa, Ontario: Canadian Health Services Research Foundation.
4. Canadian Health Services Research Foundation (January, 2006). *Weighing up the evidence: Making evidence-informed guidance accurate, achievable, and acceptable. A summary of the workshop held on September 29, 2005*. Ottawa, Ontario: Canadian Health Services Research Foundation.
5. Substance Abuse and Mental Health Services Administration—National Registry of Evidence Based Programs and Practices. (2008). Retrieved July 18, 2011 from <http://www.nrepp.samhsa.gov/Search.aspx>.
6. Institute of Medicine—Committee on Quality of Health Care in America. (2001). *Crossing the quality chasm: A new health system for the 21st century*. Washington, D.C.: National Academy Press.
7. Brownson, R. C., Fielding, J. E., & Maylahn, . M. (2009). Evidence-based public health: A fundamental concept for public health practice. *Annual Review of Public Health, 30*, 175-201.
8. Orleans, Gruman, & Anderson, 1999 (March 4, 1999). Roadmaps for the next frontier: Getting evidence-based behavioral medicine into practice. Paper presented at Society of Behavioral Medicine, San Diego, CA.
9. Victora, C., Habicht, J. P., & Bryce, J. (2004). Evidence-based public health: Moving beyond randomized trials. *American Journal of Public Health, 94*, 400-405.
10. Mihalic, S., Irwin, K., Elliott, D., Fagan, A., & Hansen, D. (2001). *Blueprints for Violence Prevention* (NCJ 187079). Washington DC: U.S. Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention.
11. Winkle, E., Luckie, A., Gehl, M., & Norris, D. (2008). *Integrating evidence-based practices into CBCAP programs: A tool for critical discussions*. FRIENDS National Resource Center for Community-Based Child Abuse Prevention.
12. Chadwick Center for Children and Families. (2004). *Closing the quality chasm in child abuse treatment: Identifying and disseminating best practices*. San Diego, CA.

13. Doll, L., Bonzo, S., Mercy, J., Sleet, D., & Haas, E (eds.). (2007). *Handbook of Injury and Violence Prevention*. Springer: New York.
14. The Community Guide to Preventive Services. *Systematic Review Methods*. Retrieved on July 18, 2011 from <http://www.thecommunityguide.org/about/methods.html>.
15. California Evidence-Based Clearinghouse. *Scientific Rating Scale*. Retrieved on July 18, 2011 from <http://www.cebc4cw.org/ratings/scientific-rating-scale/>.
16. What Works Clearinghouse. Retrieved on July 18, 2011 from <http://ies.ed.gov/ncee/wwc/>.
17. Find Youth Info. Retrieved on July 18, 2011 from <http://www.findyouthinfo.gov/>.
18. Promising Practices Network for Children, Families, and Communities. Retrieved July 18, 2011 from <http://www.promisingpractices.net/criteria.asp#evidence>.
19. World Health Organization (2009). *Violence Prevention: the evidence*. WHO Press. Geneva, Switzerland.
20. US Department of Health and Human Services (June, 2005). *Code of federal regulations: Protection of human subjects*. Retrieved July 18, 2011 from <http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.html>.
21. Food and Drug Administration (April, 2008). *Code of federal regulations: Protection of human subjects*. Retrieved July 18, 2010 from <http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=50&show>.
22. U.S. Department of Health and Human Services. Centers for Disease Control and Prevention. (2005). *Introduction to program evaluation for public health programs: A self-study guide*. Atlanta, GA: Centers for Disease Control and Prevention.
23. Hoyle, R., Harris, M., & Judd, C. (2001). *Research Methods in Social Relations (7th Ed)*. New York: Wadsworth Publishing.
24. Brewer, M. (2000). Research Design and Issues of Validity. In Reis, H. and Judd, C. (eds.) *Handbook of Research Methods in Social and Personality Psychology*. Cambridge: Cambridge University Press.
25. "Meta-Analysis." *Encyclopedia of Public Health*. Ed. Lester Breslow. Gale Cengage, 2002. eNotes.com. 2006. Retrieved on July 18, 2011 from <http://www.enotes.com/public-health-encyclopedia/meta-analysis>.
26. Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M. & Wallace, F. (2005). *Implementation Research: A Synthesis of the Literature*. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231).

For more information, please contact Centers for Disease Control and Prevention

4770 Buford Highway, MS F-64, Atlanta, GA 30341

Telephone: 1-800-CDC-INFO (232-4636)/TTY: 1-888-232-6348

E-mail: cdcinfo@cdc.gov

Web: www.cdc.gov/violenceprevention

National Center for Injury Prevention and Control
Division of Violence Prevention

