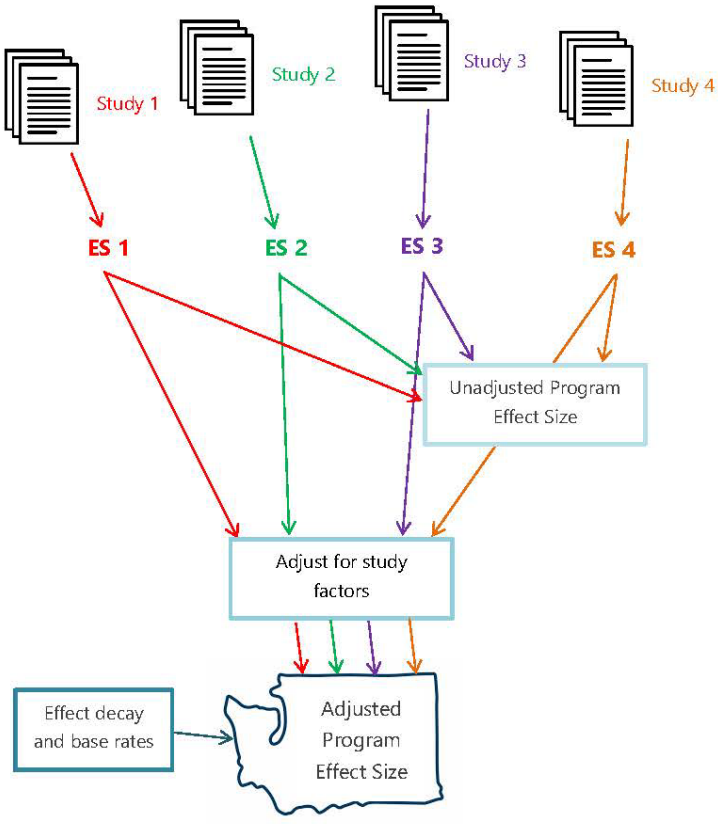
Estimating Program Effects Using Effect Sizes

A Brief Guide

WSIPP's goal is to develop practical information for the Washington State Legislature on what would happen if a program were implemented in Washington. To do this, WSIPP reviews research and summarizes information on the effects of various programs on outcomes of legislative or policy interest. WSIPP calculates a statistic for each outcome—an effect size (ES). WSIPP uses meta-analysis to create a program average effect size. This program effect size represents the average effect of the program as measured in high-quality studies. Using both the program effect size and other key information, WSIPP estimates the size of the expected change in Washington State over time if a program were implemented. The below graphic shows the steps involved in this process, and the text provides details on each step.

Overview of WSIPPs' Meta-Analysis Process

1. Gather studies which measure the impact of a program on an outcome of interest.
2. Create comparability in studies' measures using an effect size (ES).
3. Use meta-analysis to create a program average effect size.
4. Use additional information about studies to adjust the effect size.
5. Project the effect of the program in Washington over time.



1 Gather Studies which Measure the Impact of a Program on an Outcome of Interest

The first step is to determine the scope of the analysis, including the outcome of interest and target population. For example, early childhood education programs that aim to raise academic achievement as measured by outcomes such as test scores or high school graduation. Such programs can be universally available or specifically for certain students (e.g., students from low-income households). Moreover, they can be recognized name-brand programs (e.g., Head Start), or collections of similar non name-brand programs, (e.g., state and district-run early education programs).

WSIPP researchers conduct a thorough review of the research literature to find studies that evaluate the effect of these programs. Many of these studies are published in peer-reviewed academic journals while others are from sources such as government agencies or independent evaluation contractors. These studies measure the effects of programs on various outcomes—the measurable changes in results such as high school graduation or illicit drug use that occur as a result of a program. For a study to be included in WSIPP's analysis, it must be conducted using methods that allow researchers to conclude that the program caused the measured changes.

2 Create Comparability in Studies' Measures Using an Effect Size

For each rigorous study, WSIPP researchers code key characteristics and data about the study and about each measured outcome. WSIPP uses effect sizes to standardize the measurements of the effects of programs so that they can be compared on an "apples-to-apples" basis. For example, effect sizes can allow outcomes that were measured on different scales (continuous [e.g. 1-10 scale] or dichotomous [e.g., yes or no] to be directly compared and combined).[1]

An effect size is a measure of the effect of a program on a particular outcome and indicates the magnitude and direction of change. If the effect size is positive, the outcome increases. If the effect size is negative, the outcome decreases. For context, among the hundreds of effect sizes measured by WSIPP, the magnitude nearly always falls between -2.0 and 2.0 and over half fall between -0.2 and 0.2. However, effect sizes are dependent on the context where they were measured and should not be directly compared without additional context. For more information see Section 2.3 of WSIPP's Technical Documentation.

[1] In certain instances, the effect size is not the appropriate measure of program effectiveness and WSIPP conducts a meta-analysis using a different, standard measure for that literature. This occurs for incident rate ratios in the measurement of falls, percent change in the measurement of earnings and total health care costs, and elasticities in the measurement of crime rates in the policing and incarceration literature. More information is available in the Technical Documentation.

3 Use Meta-Analysis to Create a Program Average Effect Size

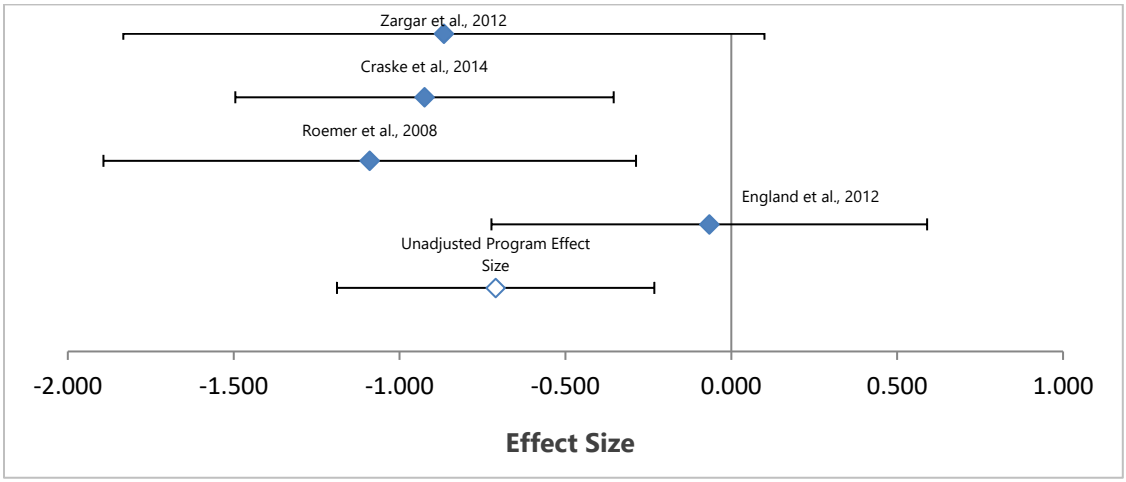
WSIPP creates an average effect size for each outcome of a program using meta-analysis. Meta-analysis is a statistical technique that creates a weighted average of the observed effects from multiple studies.[2]

The forest plot below displays the meta-analytic process. The plot shows the effect sizes for changes in anxiety from studies on the effect of "Acceptance and Commitment Therapy (ACT) for Anxiety (Adult)." The further a diamond is from 0 (the vertical line), the greater the measured impact. In this example, effect sizes less than zero represent a decrease in the level of anxiety. The lines extending from each diamond represent a possible range of the effect size based on the information provided in the study.

The white diamond at the bottom is the unadjusted program effect size—the weighted average of the effect sizes from the studies.

Exhibit 1

Forest Plot of Effect Sizes for "Acceptance and Commitment Therapy (ACT) for Anxiety (Adult)"



[2] WSIPP uses inverse variance weights. More information can be found in Section 2.3e of the Technical Documentation.

4 Use Additional Information about Studies to Adjust the Effect Size

As discussed earlier, WSIPP's goal is to supply the Washington State Legislature with information about what works to improve outcomes in Washington. WSIPP projects the effect of programs in Washington by making adjustments to the effect sizes to account for the following:

a) The methodological quality of each study we include in the meta-analyses;

b) The degree to which findings for a particular sample of people can be generalized to other populations in Washington; and

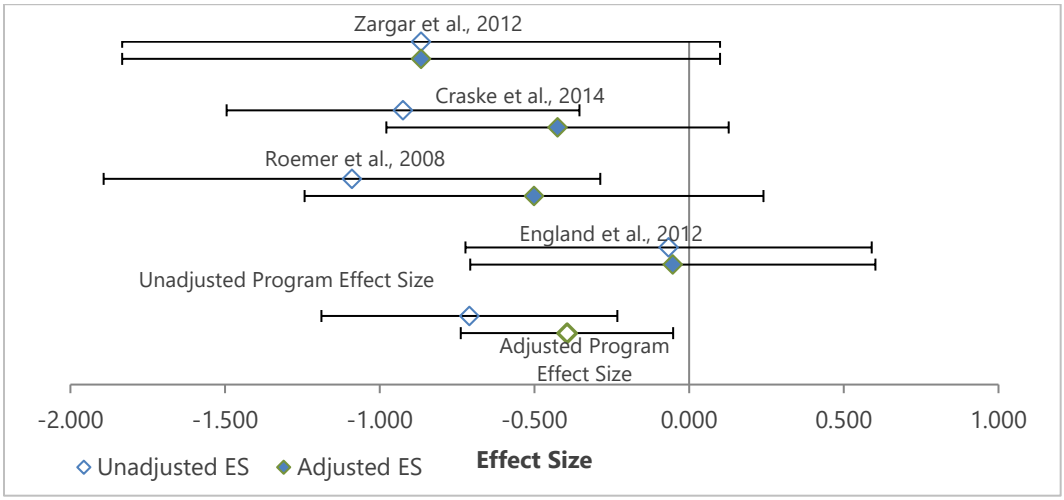
c) The relevance of the independent and dependent measures that individual studies examined.

These adjustments help create a realistic projection of what the expected effect would be in Washington State.

For example, in the scenario shown in Exhibit 1, the meta-analysis program average effect is -0.710. However, in one of the studies, the program providers were authors on the paper. WSIPP's analysis across hundreds of studies in the area of adult mental health indicates that, for programs targeting adult anxiety, studies conducted by the program provider had consistently larger effect sizes than studies conducted in environments where the program developer was not involved. If the program were to be implemented in Washington, the program developer would not be expected to be involved. Given that, WSIPP adjusts the expected size of the study by X as detailed in WSIPP's Technical Documentation. This adjustment reduces the overall effect as shown in the table below. WSIPP applied a similar adjustment to two of the studies that used a wait-list design. After applying these adjustments, the adjusted program effect size is -0.395.

Exhibit 2

Forest Plot with Adjusted Effect Sizes for "Acceptance and Commitment Therapy (ACT) for Anxiety (Adult)"



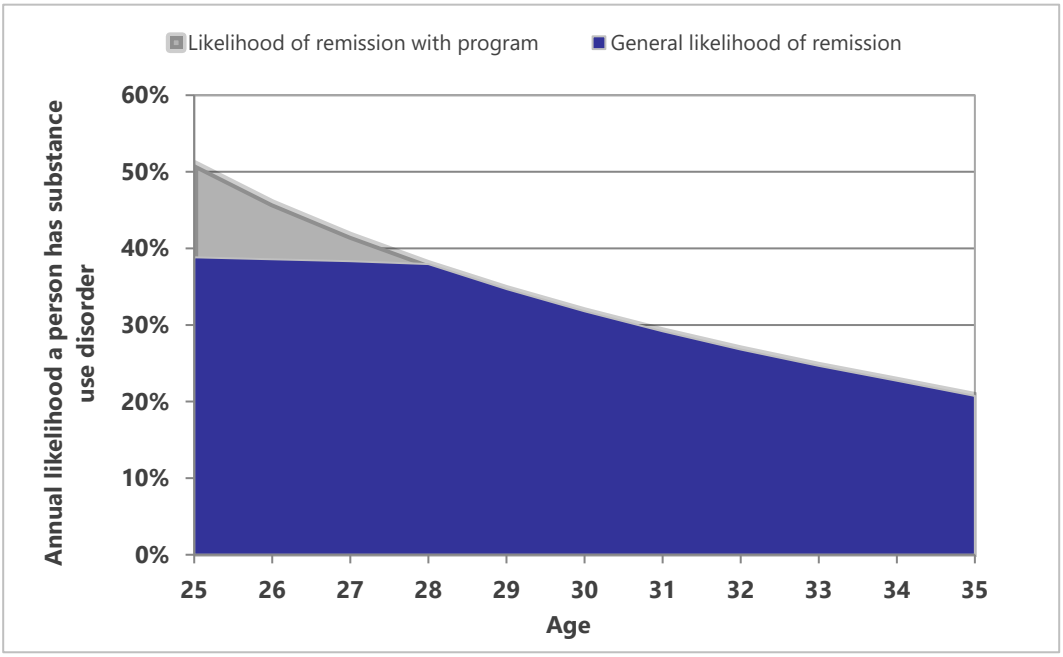
Three of the study effect sizes are reduced in magnitude to reflect that the conditions of the study do not match those in Washington. As a result, the unadjusted program effect size is replaced by a new weighted average—the adjusted program effect size.

5 Project the Effect of the Program Over Time

WSIPP recognizes that the effects of a program may last many years, so WSIPP's estimates reflect the total effect of the program over the life course. Since programs are often measured only a few years after they end, WSIPP uses available evidence from studies to project how an outcome remains (persists) or decreases (decays) over time. WSIPP looks at measurements of the outcome at different points in time and at studies specifically designed to measure the persistence of changes in the outcome. WSIPP uses that information to estimate the effect in the future (labeled as "second time ES is estimated"). The second ES estimation determines at what level the program effects are projected to continue into the future. Some effects to one-time events, such as high school graduation, do not change. Other effects, such as remission from illicit drug use disorder—shown in Exhibit 3—may fade as people who received the program relapse or those who did not receive treatment experience remission from substance use disorder—the difference between the two groups shrinks over time. For programs where the proportion of people with a particular outcome (e.g., substance use disorder) varies at different points in the life course, the projected change is the difference between those who receive the treatment and those who do not.

Exhibit 3

Example of the Expected Change in Effects over Time

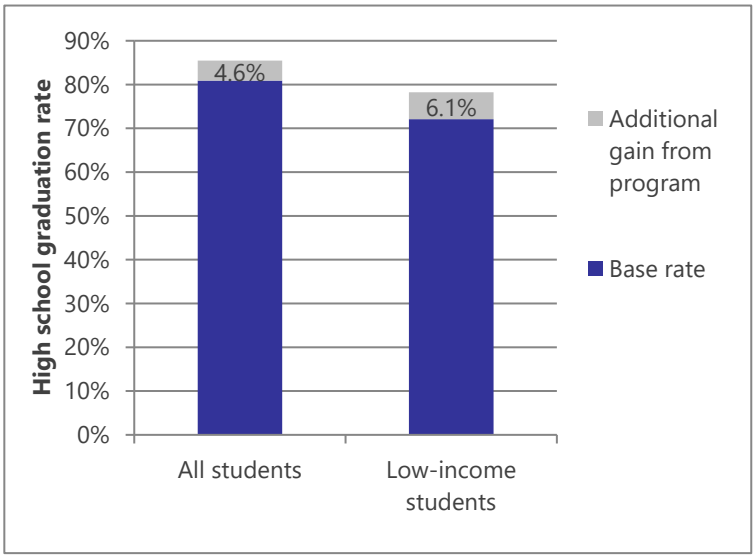


WSIPP's benefit-cost model applies the projected change to a "base rate," a measure of the current level of activity in the population that would receive the program in Washington. Programs may target specific populations for an intervention, such as children from low-income families, individuals at high risk of criminal recidivism, or individuals with a particular substance use or mental health disorder. WSIPP models these populations in Washington to create a base rate—the level of the outcome in Washington without the intervention. For example, WSIPP would look at the high school graduation rate of students from low-income households when looking at a program targeted towards those households. Additional information on those populations can be found in the Technical Documentation. The specific populations selected for each program can be found on individual program pages.

The base rate affects the size of the estimated monetizable change. This is intuitive in continuous outcomes such as test scores, where populations with a lower and greater spread of scores have more possibility of change. For example, Exhibit 4 shows that when applying the same effect size, the expected change in high school graduation rates for low-income students is much greater than for the general population.[3]

Exhibit 4

Example of the Differences in Expected Change for Different Base Rates



At the end of this process, WSIPP has estimated the expected change to an outcome from the program in Washington. WSIPP uses the change to calculate the program's benefits using the WSIPP benefit-cost model. If a program has effects on multiple outcomes, WSIPP follows these steps for each outcome.

[3] Since the effect size is also tied to the population in which it was measured, WSIPP ensures the base rate is appropriate for the literature where it was measured.