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, we review research and summarize how programs affect outcomes of legislative or policy interest. For each outcome, we calculate a statistic known as the effect size (ES) using a methodological technique called meta-analysis. This effect size represents the program's average effect measured across available academic studies. WSIPP historically adjusted some effect sizes based on the methodological quality of studies or various program characteristics. For programs reviewed in 2024 and moving forward, we will no longer calculate an adjusted effect size but will instead account for these factors in our study inclusion decisions. However, as we transition to the new methodology, our website will contain programs evaluated using both methods, so we keep this here as a reference.

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. (Pre-2024) 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, we may examine how early childhood education programs raise academic achievement as measured by outcomes such as test scores or high school graduation. Such programs can be universally available or for a limited group of students (e.g., students from low-income households). Moreover, they can be recognized as "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 thoroughly review the literature to find studies evaluating 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, resulting from participation in 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.

Beginning in 2024, we further eliminate studies that would lead to results that would not replicate in Washington. For example, we may remove studies of medium quality, those that study populations that do not apply to Washington. Historically, we included those studies but then adjusted for them in step 4 below. Due to improvements in our methodologies, we no longer make these adjustments.

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 each measured outcome. WSIPP uses effect sizes to standardize the measurements of the effects of programs so that outcomes can be compared on an "apples-to-apples" basis. For example, effect sizes can allow outcomes measured on different scales to be directly compared and combined. For example, studies that examine a continuous outcome, like a 1-10 scale, and studies that examine a dichotomous outcome, like a yes or no outcome, can be combined using this technique.[1]

[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.

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 in over half, the magnitude falls between -0.2 and 0.2. However, effect sizes depend 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.

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

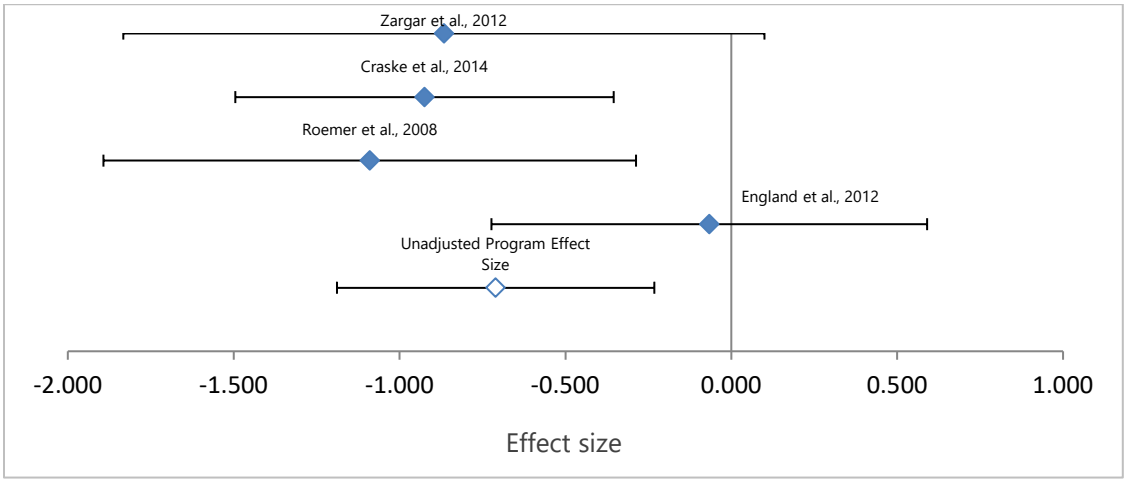
WSIPP creates an average effect size for each program outcome using meta-analysis. Meta-analysis is a statistical technique that creates a weighted average of the observed effects from multiple studies.[2]

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 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 Document.

4 (Pre-2024) Adjust the Effect Size

Historically, we adjusted the effect sizes to account for methodological quality and other factors.[3]

For example, in the scenario shown in Exhibit 1, the meta-analysis program's average effect is -0.710. However, in one of the studies, the program providers were the paper's authors, which has been associated with larger effect sizes than we would expect in Washington. Given that, WSIPP adjusted the study's effect size, reducing the overall effect. WSIPP applied a similar adjustment to two studies that used a wait-list design. After applying these adjustments, the adjusted program effect size is -0.395.

Beginning in 2024, we no longer make these adjustments but instead account for these factors in Step 1.

5 Project the Effect of the Program Over Time

The program's effects may last many years, so WSIPP's estimates reflect the program's total impact over a participant's lifetime. Since programs are often measured only a few years after they end, we use available evidence from studies to project how an outcome remains (persists) or decreases (decays) over time. We call this the "second time ES is estimated," and this estimate determines how the program effects are projected to continue. The effect size cannot change over time for one-time events, such as high school graduation. Other effects, such as remission from illicit drug use disorder, may fade as people who received the program relapse or those who did not receive treatment experience remission from substance use disorder.

WSIPP's benefit-cost model applies the projected change from program participation to a "base rate," a measure of the current activity level in the population that would receive the program in Washington. These may differ by the population studied. 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.

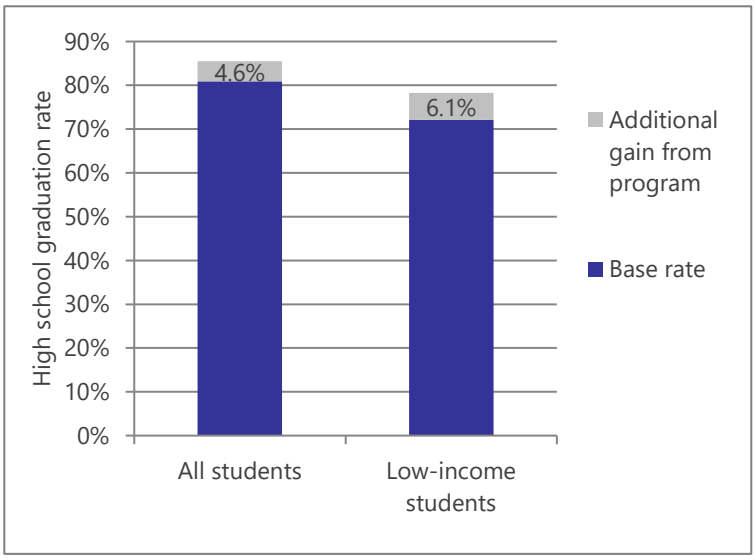
The base rate affects the size of the estimated monetizable change. This is intuitive in continuous outcomes such as test scores, where populations with a lower and greater spread of scores have more possibility of change. For example, Exhibit 2 shows that when applying the

[3] Further information about pre-2024 adjustments can be found in the 2023 Technical Document.

same effect size, the expected change in high school graduation rates for low-income students is greater than for the general population.

Exhibit 2

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 affects multiple outcomes, WSIPP follows these steps for each outcome.

For further information, contact:
Heather Grob at 360.664.9081, heather.grob@wsipp.wa.gov