PATHWAYS TO WORK
Evidence Clearinghouse

Protocol for the Pathways to Work Evidence Clearinghouse

Methods and Standards, Version 2.0 - DRAFT

This page intentionally left blank.

Acknowledgements

Many people contributed in significant ways to this report. First, we acknowledge the valued support of staff at the Administration for Children and Families, U.S. Department of Health and Human Services. We particularly thank our project officers, Marie Lawrence, Amelia Popham, and Siri Warkentien, and project monitor, Clare DiSalvo (VPD Government Solutions), for their oversight and guidance throughout the project.

We acknowledge the support we received from colleagues at Abt Global (Abt) and the American Institutes for Research (AIR). Karin Martinson (Abt) and Samia Amin (AIR) serve as co-Principal Investigators and provided guidance throughout the project, including development of the revised evidence standards and new cost standards. Ryan Williams and Sarah Sahni (AIR) provided valuable input on the standards, as did Andrew Clarkwest (Abt). Supriya Tamang, Megan Robinson, and Eleanor Elmudesi (Abt) provided dedicated, thoughtful support as we convened the panel listed below.

We would also like to thank the following members of the Pathways to Work Evidence Clearinghouse Evidence Review Expert Consultation Panel and Cost Expert Advisory Panel who provided input on the methods and standards for the review. The views expressed in this publication do not necessarily reflect the views of these members.

Jon Baron
Coalition for Evidence-Based Policy

Clive Belfield
City University of New York

Brooks Bowden
University of Pennsylvania

Catherine Darrow
J-PAL North America

Kinsey Dinan
New York City Department of Social Services

Mary Farrell
MEF Associates

Karen Gardiner
Urban Institute

Lynn Karoly
RAND

Stephanie Lee
Washington State Institute of Public Policy

David Long
Princeton Policy Associates

Cynthia Miller
MDRC

Maggie Renno
Wisconsin Department of Children and Families

Dana Rotz
Wizards of the Coast

Louisa Treskon
MDRC

Herb Turner
Analytica, Inc.

Overview

Over the past several decades, evaluations have produced a great deal of research on programs designed to improve the employment outcomes of Temporary Assistance for Needy Families (TANF) recipients, beneficiaries of other public benefit programs, and other people with low incomes. However, the sheer volume of research combined with the diversity of the findings can make it challenging to find applicable research, identify the most reliable and relevant studies, and use information from the studies to inform practical decision making. Using this research to guide decision making is complicated by the nature of programs and policies that aim to improve employment, earnings, and other related outcomes, which often involve multiple components and lead to changes in outcomes through complex pathways.

To help decision makers use this research literature, and in response to the Consolidated Appropriations Act of 2017 (Pub. L. 115-31), the Office of Planning, Research, and Evaluation established the Pathways to Work Evidence Clearinghouse. The Pathways to Work Evidence Clearinghouse seeks to be a comprehensive resource that a range of audiences, including state and local TANF administrators, can use to identify the services that will best help people with low incomes succeed in the labor market. To become this comprehensive resource, the Pathways to Work Evidence Clearinghouse aims to answer the following research questions:

1. What research exists on the effectiveness of programs and policies that have the primary aim of improving the employment and earnings of people with low incomes?
2. Which programs and policies have evidence of improving employment, earnings, education, and training for people with low incomes and of reducing public benefit receipt?

This report provides a revised protocol for the Pathways to Work Evidence Clearinghouse, describing the methods and standards used by the clearinghouse team to answer these questions. Key revisions made to the protocol include:

- Updates to the literature search protocol to ensure comprehensiveness (Section 2.1).
- Revisions to baseline equivalence standards to permit use of baseline effect sizes to establish equivalence and to waive a requirement for earnings to be measured at least one year prior to the start of the intervention under certain conditions (Section 3.2.5).
- Updates to service tags used to describe and classify programs (Section 4.1).
- New standards for including and rating the quality of program cost studies (Chapter 5).

We first detail how the Pathways to Work Evidence Clearinghouse team identifies and prioritizes eligible studies for review. Next, we document how the team determines the quality of evidence provided by eligible studies, including how team members assess, document, and assign quality ratings to studies. We describe how team members assign effectiveness ratings to these programs based on the results of the reviews and the information documented. Finally, we describe standards used to rate the quality of program cost studies.

Contents

Acknowledgements ..... **ii**

Overview **iii**

1. Introduction **3**
1.1. Key Terms 4
1.2. Overview of Review Process 5

2. Identifying Eligible Studies **7**
2.1. Searching the Literature 7
2.1.1 Integrate Studies Considered by ESER 7
2.1.2 Examine Existing Literature Reviews 7
2.1.3 Search Electronic Citation Databases 7
2.1.4 Search of Organizational Websites 8
2.1.5 Coordinate With Other Federal Evidence Reviews 9
2.1.6 Issue Calls for Papers 9
2.2. Screening Research Against Eligibility Criteria 9
2.3. Prioritizing Research for Review 11

3. Assessing A Study's Strength Of Evidence **12**
3.1. Identifying Studies Within and Across Manuscripts 12
3.2. Standards for Reviewing Studies 13
3.2.1 Study Quality Ratings 13
3.2.2 Outcomes Eligible for Review 13
3.2.3 Determination of Study Design 17
3.2.4 Standards for Reviewing RCTs 18
3.2.5 Standards for Reviewing Comparison-Group QEDs 20
3.3. The Review Process 22
3.3.1 Study Review Process 23
3.3.2 Challenges to Review Findings 23
3.3.3 Handling Potential Conflicts of Interest 23
3.4. Extracting Data From a Study 24

4. Assessing the Evidence of Effectiveness for a Program **27**
4.1. Grouping Studies Into Programs 27
4.2. Determining Programs' Evidence of Effectiveness 28
4.3. Presenting Findings 31

5. Assessing Cost Information For A Program **32**
5.1. Threshold Standards for Cost Analyses 33
5.2. Quality Rating Standards for Cost Analyses 34

References **38**

Appendix A: Process for Developing and Revising Review Protocol **A-1**

Appendix B: PRISMA-P and PRISMA-CI Elements **B-1**

Appendix C: Attrition Boundary **C-1**

Appendix D: Details on Effect Size Calculation **D-1**

Appendix E: Cost-Effectiveness and Cost-Benefit Analysis Standards **E-1**
E.1 Cost-Effectiveness Analysis Standards E-1
E.1.1 Cost-Effectiveness Analysis Threshold Standards E-1

E.1.2 Cost-Effectiveness Analysis Quality Rating Standards E-2
E.2 Cost-Benefit Analysis Standards E-5
E.2.1 Cost-Benefit Analysis Threshold Standards E-5
E.2.2 Cost-Benefit Analysis Quality Rating Standards E-7
Appendix F: Glossary of Cost Study Terms F-1

List of Exhibits

Exhibit 1.1. Research targeted by the Pathways to Work Evidence Clearinghouse review 4
Exhibit 1.2. Example of how studies, findings, and manuscripts may be nested within a program 4
Exhibit 1.3. Overview of the Pathways to Work Evidence Clearinghouse review effort 6
Exhibit 2.1. Database search strategy for the Pathways to Work Evidence Clearinghouse 8
Exhibit 3.1. Study Quality Ratings 13
Exhibit 3.2. Selecting findings for review, by outcome domain and measure 15
Exhibit 3.3. Process for reviewing RCTs 18
Exhibit 3.4. Process for reviewing comparison-group QEDs 20
Exhibit 3.5. Key data elements collected during study review process 25
Exhibit 4.1. Potential effectiveness ratings of programs, by domain 29
Exhibit 4.2. Potential designations of a program with two studies 30
Exhibit 5.1. Process for reviewing cost studies 32
Exhibit B.1. PRISMA-P Elements B-1
Exhibit B.2. PRISMA-CI methods elements not discussed in PRISMA-P B-2
Exhibit C.1. Highest differential attrition rate for sample to maintain low attrition, by overall attrition rate C-1
Exhibit D.1. Alternative estimates of Hedges' *g* D-2

1. Introduction

Over the past several decades, evaluations have produced a great deal of research on programs designed to improve the employment outcomes of Temporary Assistance for Needy Families (TANF) recipients, beneficiaries of other public benefit programs, and other people with low incomes. However, the sheer volume of research combined with the diversity of the findings can make it challenging to find applicable research, identify the most reliable and relevant studies, and use information from the studies to inform practical decision making. Using this research to guide decision making is complicated further by the nature of programs and policies that aim to improve employment, earnings, and other related outcomes. These programs and policies typically involve multiple components and lead to changes in outcomes through complex pathways (Guise et al., 2017a; Guise et al., 2017b).

To help decision makers use this research literature, and in response to the Consolidated Appropriations Act of 2017 (Pub. L. 115-31), the Office of Planning, Research, and Evaluation (OPRE), within the Administration for Children and Families (ACF), U.S. Department of Health and Human Services established the Pathways to Work Evidence Clearinghouse. The Pathways to Work Evidence Clearinghouse seeks to be a comprehensive resource that a range of audiences, including state and local TANF administrators, can use to identify the services that will best help people with low incomes succeed in the labor market.

This report describes the methods and standards used to conduct reviews for the Pathways to Work Evidence Clearinghouse. The Pathways to Work Evidence Clearinghouse seeks to provide a systematic assessment of the effectiveness of employment-related services and policies for people with low incomes (Moher et al., 2015; Munn et al., 2018).[1] To provide this assessment, the Pathways to Work Evidence Clearinghouse set out to answer the following research questions:

1. What research exists on the effectiveness of programs and policies that have the primary aim of improving the employment and earnings of people with low incomes?

2. Which programs and policies have evidence of improving employment, earnings, education, and training for people with low incomes and of reducing public benefit receipt?

Many systematic reviews of evidence apply the PICOTS (population, intervention, comparators, outcomes, timing, and setting) framework to explain the review's scope (Thompson et al., 2012). Exhibit 1.1 formally summarizes the scope of this review, using that framework.

[1] The reviews conducted under the Pathways to Work Evidence Clearinghouse comprise a scoping review, rather than a systematic review, as its goal is to use systematic methods for searching, selecting, and synthesizing research to understand the broad field of evidence on employment and training programs for people with low incomes, rather than using these methods in a targeted review that solely assesses the evidence on specific programs (Colquhoun et al., 2014).

Exhibit 1.1. Research targeted by the Pathways to Work Evidence Clearinghouse review

Population	People ages 16 and older with low incomes
Interventions	Programs, policies, and strategies with a primary aim of improving employment and earnings
Comparators[a]	Services typically provided to people with low incomes or other programs and policies for which people with low incomes might be eligible
Outcomes	Employment, earnings, public benefit receipt, and attainment of education and training credentials
Timing	The review restricts attention to analyses conducted in 1990 and later. The services or policies implemented within any particular intervention can be of any duration.
Setting	United States and Canada

Note: Classification based on PICOTS framework. See Thompson et al. (2012).
[a] In the PICOTS framework, these are services provided to the comparison group in the targeted research.

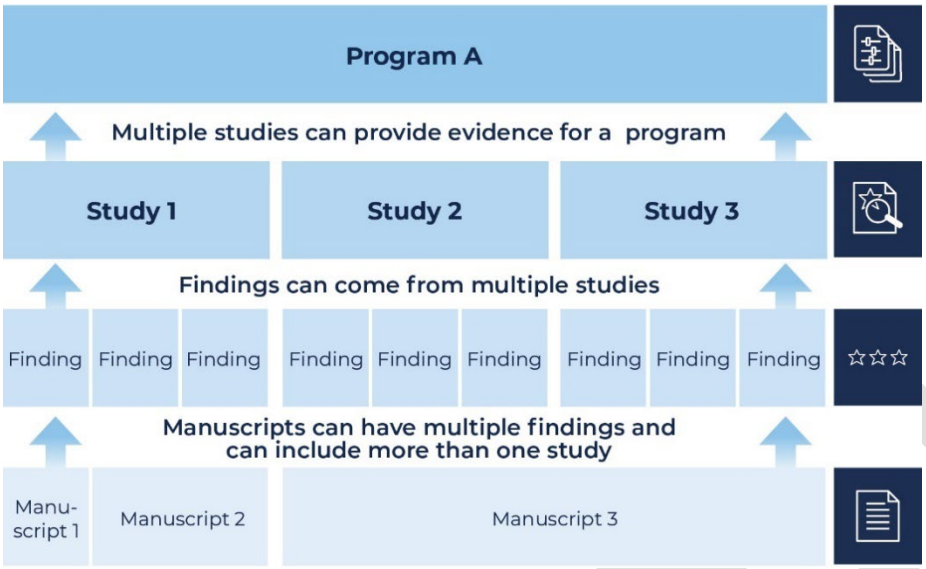
1.1. Key Terms

The Pathways to Work Evidence Clearinghouse relies on specific terminology to classify research. It defines the following terms.

- A **program** is a specific bundle of services or policies implemented in a given context.
- A **study** is an analysis of a distinct implementation of a program.
- **Findings** summarize the effect of a program on an **outcome** measure related to employment, earnings, public benefit receipt, education, or training.
- **Manuscripts**, which describe studies, may include published and unpublished research, such as journal articles, working papers, and book chapters. Note that in some cases, one manuscript may include several studies; in other cases, one study may be reported across multiple manuscripts.

Exhibit 1.2 provides an example of how each term might relate to a single program. For more detail about how the Pathways to Work Evidence Clearinghouse groups research into studies, see Section 3.1.

Exhibit 1.2. Example of how studies, findings, and manuscripts may be nested within a program



1.2. Overview of Review Process

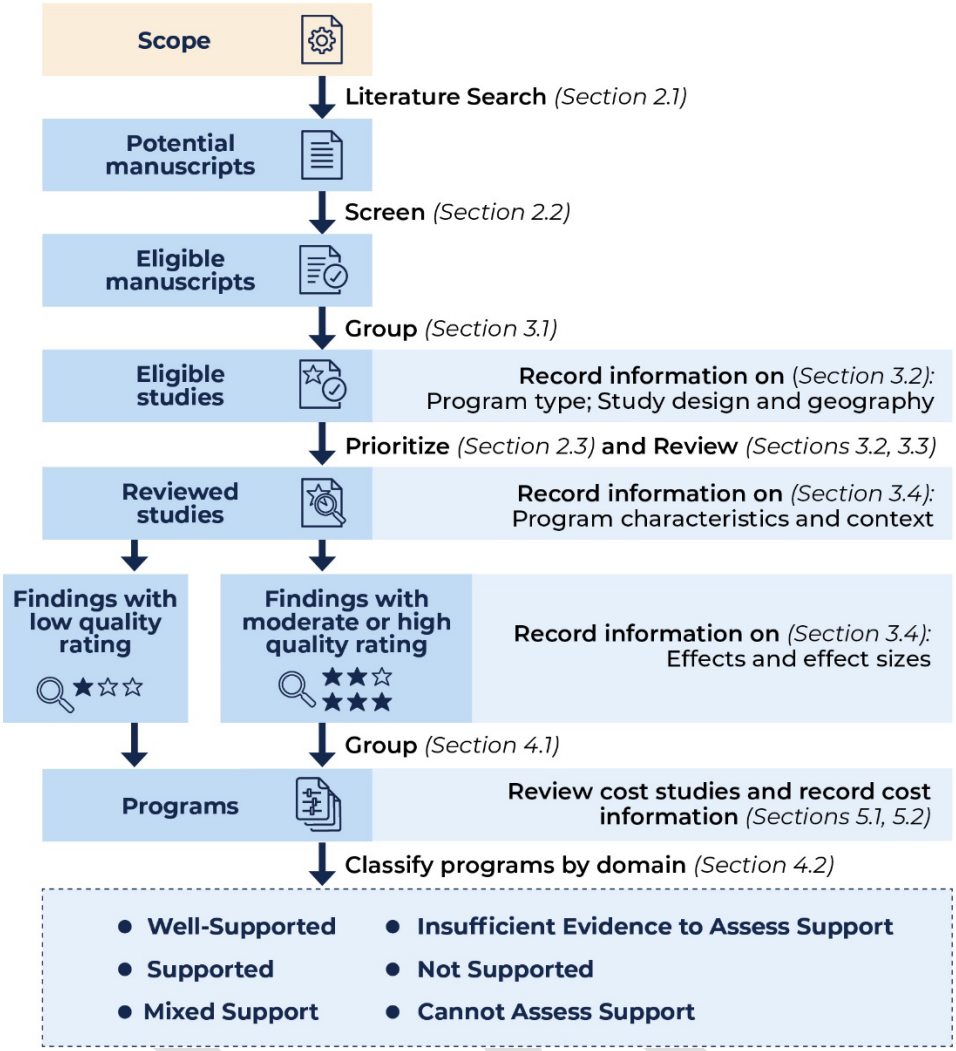
The remainder of this report provides a protocol that is used for each review cycle conducted, describing the overall approach to accomplishing the goals of the Pathways to Work Evidence Clearinghouse (Exhibit 1.3 below also provides a graphical overview).[2] In Chapter 2, we detail the scope of the review, how the Pathways to Work Evidence Clearinghouse team searches for manuscripts containing eligible research, and how it prioritizes research for review. In Chapter 3, we document how studies are identified within and across manuscripts and how reviewers assess, document, and assign quality ratings to studies. In Chapter 4, we describe how staff assign effectiveness ratings to programs based on the results of the study reviews and the information documented from one or more studies of a program. In Chapter 5, we present cost standards designed to assess the quality of cost estimates. Appendix A provides a brief overview of the process used to develop and revise this protocol.

The review plans specified here are aligned with each section of the Preferred Reporting Items for Systematic Reviews and Meta-Analysis for Protocols (PRISMA-P; Moher et al., 2015) and the methods section of the PRISMA for Complex Interventions (PRISMA-CI; Guise et al., 2017b). This report also serves as the protocol for the review (PRISMA-CI element 5). A checklist version of the PRISMA-P and PRISMA-CI elements appears in Appendix B.

If the scope of the Pathways to Work Evidence Clearinghouse is updated, or any other notable changes are made to its protocol, standards, or procedures, we will issue a revised version of this report.

[2] The project uses a pair of databases to catalog manuscripts and their corresponding studies as a management tool to track the literature search, screening, and review processes.

Exhibit 1.3. Overview of the Pathways to Work Evidence Clearinghouse review effort



2. Identifying Eligible Studies

This chapter details the process for identifying eligible studies to include within the Pathways to Work Evidence Clearinghouse. Section 2.1 describes the process for searching the literature, Section 2.2 details the screening process for eligibility, and Section 2.3 provides an overview of how research is prioritized for review.

2.1. Searching the Literature

Team members use the following six strategies to identify potential research for review.

2.1.1 Integrate Studies Considered by ESER

The team included all research reviewed by an OPRE-funded precursor project, the Employment Strategies for Low-Income Adults Evidence Review (ESER) in the Pathways to Work Evidence Clearinghouse review. As the scope of the Pathways to Work Evidence Clearinghouse has evolved (see Appendix A), the team also re-examined manuscripts that had been identified during the ESER literature search but were considered "out of scope" for ESER (and, thus, not reviewed by ESER) to see if they met the revised eligibility criteria (see Section 2.2 below).

2.1.2 Examine Existing Literature Reviews

The Pathways to Work Evidence Clearinghouse team scans literature reviews on an ongoing basis to identify any new literature reviews on the effectiveness of employment and training programs. Relevant literature reviews are identified through multiple sources, including: (a) regular consultation with internal content experts, (b) literature reviews and meta-analyses cited in studies reviewed, (c) checking citations identified from the literature search that are screened out for being a meta-analysis or literature review. A list of literature reviews consulted by the Pathways to Work Evidence Clearinghouse is maintained on the Pathways to Work website, noting the date that the list was most recently updated.

2.1.3 Search Electronic Citation Databases

The Pathways to Work Evidence Clearinghouse team systematically searches several electronic citation databases to find relevant research published since the most recently conducted search. We summarize the search strategy in Exhibit 2.1 below.

Exhibit 2.1. Database search strategy for the Pathways to Work Evidence Clearinghouse

Criterion	Keywords (Title or Abstract)
Design terms	regression OR experiment* OR quasiexperiment* OR quasi-experiment* OR pseudoexperiment* OR pseudo-experiment* OR nonexperiment* OR non-experiment* OR causa* OR statistical* OR random* OR correlat* OR descript* OR "propensity score*" OR "match* design" OR "fixed effect*" OR "fixed-effect*" OR "interrupted time series" OR "least square*" OR "treatment*" OR "intent to treat" OR "intent-to-treat" OR "instrumental variable*" OR "local average treatment effect*" OR "event stud*" OR "event history" OR (pre N2 post) OR (pre-post) OR "formative evaluation*" OR "formative stud*" OR "outcome evaluation*" OR "outcome stud*" OR "program evaluation*" OR "policy evaluation*" OR "control group*" OR "control condition*" OR "comparison group*" OR "comparison condition*" OR trial OR intervention* OR QED* OR RCT* OR "service* as usual" OR "service-as-usual" OR "services-as-usual" "usual community services" OR "usual community treatment" OR "usual services" OR "regular services" OR "wait-list" OR waitlist OR "waiting list" OR "waiting-list" OR "matched group* " OR "matched comparison"
Outcome terms	employ* OR reemploy* OR selfemploy* OR unemploy* OR wage* OR earn* OR "self-sufficien*" OR "number of jobs" OR (held n3 jobs) OR (hold n3 jobs) OR (holds n3 jobs) OR "are in work" OR "were in work" OR "is in work" OR "was in work" OR "back in work" OR "find work" OR "found work" OR "finding work" OR "returned to work" OR "return to work" OR "returns to work" OR "back to work" OR "out of work" OR "out-of-work" OR "not in work" OR "are not working" OR "were not working" OR "are working" OR "were working" OR "is working" OR "was working" OR "out-of-work" OR "career advance*" OR "job retention" OR "labor market" OR "labour market"
Sample terms	((income* OR wage* OR earning* OR skill*) N2 (low OR below OR under OR less)) OR (("public benef*" OR welfare OR AFDC OR TANF OR WIC OR SNAP OR "Food Stamp*" OR "Social Security" OR Medicaid OR "social program*" OR "social assistance" OR "income support" OR "income assistance") N3 (benefi* OR recipient* OR client* OR customer* OR participant* OR recei* OR consumer)) OR poverty OR impoverished OR FPL OR disadvantaged OR unemploy* OR underemploy* OR unskilled OR jobless OR homeless* OR houseless* OR unhoused OR housing OR "doub* up" OR "couch surf*" OR shelter* OR offender* OR exoffender* OR parole* OR probation* OR "criminal history" OR (justice n3 involve*) OR incarcerated OR (youth* N2 (disconnected OR opportunity OR "at-risk OR "at risk")) OR apprentic* OR trainee OR ((participa* OR attend* OR complet*) n3 train*) OR (poor* n3 (worker* OR individual* OR American* OR Canadian* OR citizen* OR resident* OR communit* OR member* OR famil* OR household* OR working))
Databases	Academic Search Premier, Business Source Corporate Plus, EconLit, Education Research Complete, E-Journals, ERIC, PsycINFO, Scopus, and SocINDEX with full text

Notes: The database search required a study to match at least one term for each eligibility criterion. An asterisk indicates a truncation. That is, when used as a search term, all words with the root will appear in the results. For instance, a search on "effect*" will return citations with the words that have "effect" as the first six letters, including "effect," "effects," "effective," and "effectiveness." In addition, "word1 nX word2" indicates that word1 and word2 should appear within X words of each other. For example, "holds n3 jobs" means that "holds" should appear within three words of "jobs," such as "holds more than four jobs" or "holds one job."

2.1.4 Search of Organizational Websites

The team also executes custom Google searches of key websites to identify additional studies (see Box 2.1 below). These sources of research are relevant to the review but conducting a search for them has restrictions, such as not allowing search limitations by date range or restrictions to certain fields. The team searches them in a way that matches, as closely as possible, the criteria shown in Exhibit 2.1. Finally, the team executes searches in the Harvard Think Tank Engine. This publicly available customized Google search engine searches the websites of more than 1,200 institutions that generate public policy research, analysis, and activity. These sites are affiliated with universities, governments, advocacy groups, foundations, nongovernmental organizations, and businesses.

2.1.5 Coordinate With Other Federal Evidence Reviews

The scope of the Pathways to Work Evidence Clearinghouse includes research also eligible for review by the Department of Labor's Clearinghouse for Labor Evaluation and Research (CLEAR), OPRE's Home Visiting Evidence Review of Effectiveness (HomVEE) review, and OPRE's Prevention Services Clearinghouse review. Therefore, the team examines other federal evidence review databases to identify additional studies. This effort targets the following CLEAR topic areas: Apprenticeship and Work-Based Training, Career Academies, Community College, Disability Employment Policy, Job Search Assistance, Low-Income Adults, Older Workers, Opportunities for Youth, Reemployment, Reentry (for formerly incarcerated people), and Veterans. It also targets studies in HomVEE reviews that measured family economic self-sufficiency as an outcome and studies in Prevention Services Clearinghouse reviews that include economic stability outcomes, including income and employment outcomes.

Box 2.1. Organizational websites included in custom Google search

- Abdul Latif Jameel Poverty Action Lab
- Abt Global
- Administration for Children and Families
- American Institutes for Research
- Annie E. Casey Foundation
- Arnold Ventures
- Campbell Collaboration
- Center for Law and Social Policy
- Center on Poverty and Inequality (Georgetown Law)
- Chapin Hall
- Clearinghouse for Labor Evaluation and Research
- Employment & Training Administration Research Database
- Institute for Research on Poverty (University of Wisconsin-Madison)
- IZA
- Joblessness and Urban Poverty Research Program (Harvard University)*
- Mathematica
- MDRC
- MEF Associates
- National Center for Children in Poverty (Bank Street Graduate School of Education)
- NBER Working Papers
- NORC
- Opportunity Insights
- Poverty Solutions (University of Michigan)*
- RAND
- Ray Marshall Center (University of Texas at Austin)
- RePEc
- RTI International
- Social Policy Research Associates
- Social Science Research Network
- The Stanford Center on Poverty and Inequality*
- Stone Program in Wealth Distribution, Inequality, and Social Policy (Harvard University)*
- UC Davis Center for Poverty and Inequality Research*
- University of Kentucky Center for Poverty Research
- Upjohn Institute
- Urban Institute
- William K. Kellogg Foundation

*Website of university housing the center searched

2.1.6 Issue Calls for Papers

The Pathways to Work Evidence Clearinghouse regularly issues calls for papers to encourage authors and other interested parties to share studies directly with Pathways to Work. We send these calls to research organizations, professional associations, individual researchers, expert groups, and the Pathways to Work Evidence Clearinghouse audiences, in addition to posting them on the Pathways to Work Evidence Clearinghouse website and using social media accounts, newsletters, and other distribution lists.

2.2. Screening Research Against Eligibility Criteria

To identify the research eligible for review from the list of citations identified from the literature search, trained screeners apply a set of eight criteria aligned with the scope of the Pathways to Work Evidence Clearinghouse (see Appendix A for a description of the evolution of the scope of the review). Research must meet all eight criteria to be eligible for review.

1. **Analyses must have been conducted in 1990 or later.** The Pathways to Work Evidence Clearinghouse includes research first published in 1990 or later (for unpublished manuscripts, the team will use the date the manuscript was first made available). A single study might have had multiple publications presenting its results. Publications in or after 1990 that repackage study conclusions disseminated before 1990 (such as a journal article summarizing an earlier report) are ineligible. However, publications in or after 1990 that provide new conclusions not provided in an earlier report (such as a report that provides results for a longer period than an earlier report did) are eligible.

2. **Conducted in the United States or Canada.** Eligible research examines a program implemented in the United States or Canada.

3. **Assessed effectiveness using quantitative methods.** Only studies based on randomized controlled trials (RCTs) and comparison-group quasi-experimental designs (QEDs)—including comparative interrupted time series designs—are currently eligible for review.[3] The team screens out all studies that use purely descriptive methods (for example, studies that examine only outcomes of a program and do not use a comparison group) and studies that focus only on a program's implementation.

4. **Examined a program serving people with low incomes.** Eligible research must focus on programs intended to serve people ages 16 or older at the time of enrollment with low incomes. Research meets this criterion if a manuscript's authors declare a sample or population to have low income, using any definition of low income the authors provide. Alternatively, research also meets this criterion if the authors examine a sample or population of whom the majority are in a group that the clearinghouse classifies as having low income, following ESER's approach. These include people who have experienced homelessness, people who have formerly been incarcerated, people receiving means-tested public benefits, disconnected youth, and people characterized by authors as having low skills, including those who are in adult basic education, adult literacy education, or other basic skills programs.[4] If a sample is neither in one of the specified groups nor declared to have low income, the Pathways to Work Evidence Clearinghouse considers people to have low income if all sample members have incomes below the national median income in the year the study began. The Pathways to Work Evidence Clearinghouse does not categorically classify individuals who are not currently employed to have low income because lack of employment might be temporary.

5. **Examined a program or policy aiming to directly or indirectly improve employment or earnings.** Eligible programs include employment and training programs (for example, a job search assistance or occupational training program) and programs that could indirectly improve employment through strategies such as general education (for example, a

---

[3] In some cases, comparative interrupted time series designs can be reviewed as comparison-group QEDs. The Pathways to Work Evidence Clearinghouse does not currently review regression discontinuity or instrumental variables designs, as it does not currently have standards for reviewing these designs.

[4] Research that focuses on recipients of Unemployment Insurance or Social Security Disability Insurance is not eligible for review unless recipients are explicitly classified as having low income, as these programs are not means-tested.

community college tuition assistance program), two-generation programs, and helping individuals stabilize their lives (for example, housing assistance).

6. **Examined the effect of a program or policy on employment or earnings outcomes.** Research must examine impacts on some measure of employment or earnings, including but not limited to the outcomes eligible for review described in Chapter 3, Exhibit 3.2.

7. **Examined a program or policy serving individual job-seekers in a specific context.** Research that examined policies or actions that affected communities, such as enterprise zones, or employers, such as tax credits for hiring disadvantaged workers, is not eligible for review. In addition, studies on how a policy affects an entire state or country do not meet this criterion.

8. **Articulated details on the services provided.** Because the Pathways to Work Evidence Clearinghouse aims to be a repository of information on program effectiveness, the research must describe the program examined in sufficient detail so that other studies of the same program could potentially be identified by reviewers.

The Pathways to Work Evidence Clearinghouse uses a two-stage process for study screening. In the first stage, screeners examine manuscripts' titles and abstracts and screen out duplicate citations and those that obviously do not meet the criteria for inclusion. For all studies not screened out by this initial process, screeners then skim a study's full text to finalize eligibility. Screeners record the citation and note a small number of characteristics of the program being examined and the methods used to examine it in a tracking tool created for this purpose. To ensure consistency in the screening process, all screeners receive a standard training that covers the kinds of programs eligible for review and study eligibility criteria. In addition, the screening task leader checks the disposition of each screener's initial screening of studies for accuracy.

2.3. Prioritizing Research for Review

Given work already completed under the current scope (see Appendix A), the Pathways to Work Evidence Clearinghouse will prioritize reviewing newly published research within its existing scope. Periodically, the Pathways to Work Evidence Clearinghouse team revisits the scope of the clearinghouse. The team will work with OPRE and an external team of interested parties to assess priorities and determine which expansions could be most beneficial to the field. Once a specific expansion is identified as a potential priority, the Pathways to Work Evidence Clearinghouse team will provide OPRE with an estimate of the volume of research it will likely yield and a discussion of any changes in procedures or standards that would be needed to accommodate the expansion (such as creating new standards for different research designs or relaxing eligibility criteria). If OPRE approves the expansion given this information, the Pathways to Work Evidence Clearinghouse team will then search for research within the expanded scope and screen citations for inclusion (see Sections 1 and 2 of this chapter). Within the expanded scope, recently published research will be prioritized.

3. Assessing A Study's Strength Of Evidence

Well-specified standards to assess research quality support the Pathways to Work Evidence Clearinghouse review in providing systematic and unbiased ratings of the strength of causal evidence provided by studies. This chapter provides a brief overview of how studies are identified within and across manuscripts (Section 3.1), the standards staff use to review studies (Section 3.2), the review process for applying the standards (Section 3.3), and the data extracted from studies during the review process (Section 3.4). These review standards apply to RCTs and comparison-group QEDs only.

3.1. Identifying Studies Within and Across Manuscripts

The Pathways to Work Evidence Clearinghouse reviews assess the quality of evidence at both the finding and study levels. As discussed in Chapter 1, the Pathways to Work Evidence Clearinghouse defines a study as an analysis of a distinct implementation of a program (analyses of different programs cannot constitute the same study). This means that Pathways to Work can find that an individual manuscript contains multiple studies or that a single study is presented across multiple manuscripts. How the Pathways to Work Evidence Clearinghouse identifies studies within and across manuscripts can have important implications for the conclusions of the review (see Chapter 4).

Many evaluations of employment and training programs take place in multiple locations, include individuals enrolled in the evaluation at multiple points in time, and include multiple target populations (for example, TANF applicants and TANF recipients, or men and women). The Pathways to Work Evidence Clearinghouse defers to study authors in determining whether different groups of people are subject to the same distinct implementation of a program. In particular, if study authors present analyses of employment and earnings outcomes that pool groups of people, the Pathways to Work Evidence Clearinghouse considers these people to have received the same implementation of the program. If, instead, authors only present subgroup-specific analyses for employment and earnings outcomes, the analysis of each subgroup is considered a separate study.[5, 6]

[5] If authors present both pooled and subgroup analyses, but clearly indicate that the pooled analyses should be considered supplemental or exploratory, while the subgroup analyses are the primary results of interest, the Pathways to Work Evidence Clearinghouse classifies each subgroup analysis as a separate study. In addition, if the authors present subgroup-specific analyses for employment and earnings outcomes but pooled analyses (and no subgroup-specific analyses) for education and/or public benefit receipt outcomes, the Pathways to Work Evidence Clearinghouse will review the subgroup-specific analyses of employment and earnings outcomes and the pooled analyses of education and public assistance outcomes.

[6] If authors provide separate estimates for subgroups of people defined by any characteristic other than location or time of service receipt, and those subgroups were eligible for the same services, the subgroup estimates are treated as if they are from the same study for the purposes of assigning intervention effectiveness ratings (but catalogued as separate studies on the Pathways to Work Evidence Clearinghouse website). Studies that focus on a specific site or cohort of individuals are treated as separate studies for all purposes. See Chapter 4 for details.

As an example, consider a manuscript assessing the impact of a training program implemented in three Ohio cities: Cleveland, Cincinnati, and Columbus. If the authors presented impacts both by city and combined for all individuals served in all three cities, the Pathways to Work Evidence Clearinghouse team would focus on the impacts for all individuals served and would characterize this manuscript as containing one study. If, instead, the authors only presented city-specific analyses, the Pathways to Work Evidence Clearinghouse team would treat the manuscript as if it contained three studies, one for each city. In this way, we defer to the authors' assessment of whether the implementation was similar enough in each city to warrant estimating impacts based on the sample pooled across cities.

Box 3.1. Pathways to Work Evidence Clearinghouse ratings

The Pathways to Work Evidence Clearinghouse will assign a variety of ratings to characterize findings, studies, and programs.

Findings will receive a quality rating, called the study quality rating by finding, based on the strength of the causal evidence a study provides on the effects of an intervention on the outcome associated with the finding.

Studies will receive a study quality rating based on the highest rating received by any finding from that study related to earnings, employment, public benefit receipt, or education and training outcomes.

Programs will receive effectiveness ratings within each domain (or group of outcomes, see Section 3.2.2) based on the extent to which high- and moderate-quality causal evidence indicates the program improves outcomes in that domain.

3.2. Standards for Reviewing Studies

3.2.1 Study Quality Ratings

The central goal of the study review is to assess the strength of a study's design (that is, its risk of bias) and assign it the most appropriate **study quality rating** (see Box 3.1). Exhibit 3.1 includes the possible ratings reviewers will assign to summarize a study's evidence. A **high rating** indicates that the risk is relatively low that the study produces biased estimates of a program's causal effect, while a **low rating** suggests the risk of bias could be high. A **moderate rating** falls in between: there is some risk of bias but the program is likely to have contributed to the finding to at least some extent. Reviewers first assign an individual study quality rating for each finding (related to an earnings, employment, public benefit receipt, or education and training outcome) that was selected for review. They then assign a study the highest rating given to any of its associated findings.

Exhibit 3.1. Study Quality Ratings

Rating	Interpretation
High	There is strong evidence that the findings are solely attributable to the program examined.
Moderate	There is some evidence that the findings are attributable, at least in part, to the program examined. However, other factors not accounted for in the study might also have contributed to the findings.
Low	There is little evidence that the findings are attributable, in part or as a whole, to the program examined.

3.2.2 Outcomes Eligible for Review

Many studies of employment and training programs examine a wide variety of outcome measures and include findings related to these measures at several time horizons. Including all such measures in the review could result in spurious conclusions about **statistical significance** of

findings.[7] It also might be difficult for an individual accessing the Pathways to Work Evidence Clearinghouse to sort through and make sense of study results if too many findings are included in reports.

To avoid these issues, reviewers select a limited number of findings on which to focus their reviews, using the guidance outlined in Exhibit 3.2 below. Selected findings examine outcomes falling within one of ten **domains** (or groups of related outcomes):

1. **Short-term employment:** including measures of employment status and duration and consistency of work within the first 18 months after an individual is assigned to a study group.[8]
2. **Long-term employment:** including measures of employment status and duration and consistency of work more than 18 months and up to 60 months after an individual is assigned to a study group.
3. **Very long-term employment:** including measures of employment status and duration and consistency of work more than 60 months after an individual is assigned to a study group.
4. **Short-term earnings:** including measures of earned income within the first 18 months after an individual is assigned to a study group.
5. **Long-term earnings:** including measures of earned income more than 18 and up to 60 months after an individual is assigned to a study group.
6. **Very long-term earnings:** including measures of earned income more than 60 months after an individual is assigned to a study group.
7. **Short-term public benefits:** including measures of the receipt of public benefits from programs such as TANF and the Supplemental Nutrition Assistance Program (SNAP), and the amount of benefits received, measured within the first 18 months after an individual is assigned to a study group.[9]
8. **Long-term public benefits:** including measures of the receipt of public benefits from programs such as TANF and SNAP, and the amount of benefits received, more than 18 months and up to 60 months after an individual is assigned to a study group.

---

7 The Pathways to Work Evidence Clearinghouse considers a finding to be statistically significant if the *p*-value of a two-sided hypothesis test of whether an effect is equal to zero is less than 0.05.

8 The Pathways to Work Evidence Clearinghouse defines short-, long, and very long-term based on the amount of time after an individual was first assigned to a study group. If studies instead present results based on time since services were last received, and Pathways to Work Evidence Clearinghouse reviewers cannot re-align outcomes based on the timing of service commencement (for example, because different study participants received services for different lengths of time), the team instead uses time since services were last received in differentiating between short- and long-term outcomes.

9 In order to include information on all types of benefits deemed relevant by study authors, findings related to the receipt of any benefits provided by federal, state, or local governments (other than employment and training services) are included in the public benefits domains.

9. **Very long-term public benefits:** including measures of the receipt of public benefits from programs such as TANF and SNAP, and the amount of benefits received, more than 60 months after an individual is assigned to a study group.
10. **Education and training:** including measures of the attainment of educational degrees and other credentials of potential value in the labor market.

If the research provides findings for multiple outcome measures, the team prioritizes findings for review based on outcome measure, following the prioritization process summarized in Exhibit 3.2. The finding prioritization process occurs independently for each data source.[10] For example, if both surveys and administrative records were used to assess earnings, reviewers select two sets of earnings findings for review: one measured using survey data and one measured using administrative data. The Pathways to Work Evidence Clearinghouse team selects findings based on the full sample of study participants, if available.

Reviewers also assess whether each outcome examined in a study has sufficient validity to include in the review. Many evidence reviews include specific criteria that individual outcomes must meet to be considered valid and reliable (that is, for reviewers to have confidence that the outcome correctly measures the concepts they seek to measure). The Pathways to Work Evidence Clearinghouse requires all outcomes to have **face validity**—that to a general reader, the outcome should seem to measure its intended concept. Generally, employment, earnings, public benefit receipt, and education and training outcomes are objective measures with strong face validity.

Exhibit 3.2. Selecting findings for review, by outcome domain and measure

Rules for selecting findings based on outcome measures	Notes
Employment	
Select the finding examining the outcome measure that is first in this 1. Employment status at the time of follow-up 2. Employment status during the latest available month 3. Employment status during the latest available quarter 4. Employment status during the latest available year 5. Employment status over the entire follow-up period (for example, employed since random assignment) Also select findings examining cumulative measures of employment status (such as duration of employment, quarters employed, employment over consecutive quarters, or number of consecutive time periods of employment) for the longest elapsed period (for example, 18 or 12 months for short-term, 3 years for long-term).	**Include** • One set of findings each for short-term, long-term, and very long-term, as applicable) • Findings for outcomes capturing both overall and unsubsidized employment if both are considered (or unsubsidized and subsidized employment if presented in this way) • Findings for outcomes for full- and part-time employment separately if a combined measure is not available **Exclude findings related to the following outcome** • Employment by job characteristics (for example, percentage employed in a job offering benefits) • Point in time measures of employment other than those at follow-up (for example, exclude a measure such as employed in Quarter 1)

[10] Pathways to Work Evidence Clearinghouse reviewers treat multiple rounds of data collection that leverage similar survey instruments or the same administrative database as the same data source.

Rules for selecting findings based on outcome measures	Notes
Earnings	
Select the findings examining the outcome measure that is first in this list: 1. Annual earnings for the latest elapsed year of the follow-up period 2. Average annual earnings over the follow-up period 3. Total earnings over the follow-up period 4. Quarterly earnings for the latest elapsed quarter of the follow-up period 5. Monthly earnings for the latest elapsed month of the follow-up period 6. Average hourly wage rate at follow-up 7. Median hourly wage rate at follow-up	**Include** • One set of findings each for short-term, long-term, and very long-term outcomes, as applicable) • Findings for outcomes capturing both overall and unsubsidized earnings if both are considered (or earnings from unsubsidized and subsidized employment if presented in this way) **Exclude findings related to the following outcome measures:** • Earnings measures for only people who are employed
Public benefit receipt	
Select the findings examining the outcome measures that are first in this list: 1. Indicators of benefit receipt both overall and by specific benefit type and amount (dollars) of annual benefit receipt for the longest elapsed follow-up year (for example, receipt in Year 4 of a 4-year follow-up) 2. Indicators of benefit receipt both overall and by specific benefit type and average amount of annual benefits over the follow-up period (for example, average benefits Years 1–8) 3. Indicators of benefit receipt both overall and by specific benefit type and amount of total benefits received over the follow-up period (for example, total benefits collected Years 1–3) 4. Indicators of benefit receipt both overall and by specific benefit type and amount of benefits received for the latest elapsed follow-up quarter 5. Indicators of benefit receipt both overall and by specific benefit type and amount of benefits received for the latest elapsed follow-up month	**Include** • One set of findings each for short-term, long-term, and very long-term outcomes and one set for long-term outcomes, as applicable • Findings for decompositions of benefit receipt if they are presented by study authors (for example, measures of receiving TANF, Supplemental Nutrition Assistance Program, or Unemployment Insurance benefits) • Measures of months of benefit receipt if indicators of overall benefit receipt are not available **Exclude findings related to the following outcome measures:** • Benefit amounts for only people who receive benefits
Education and training	
Select the findings examining measures of educational attainment over the follow-up period (for example, acquisition of a GED, associate's degree, bachelor's degree, attaining a certificate or credential)	**Exclude findings related to the following outcome measures:** • Decompositions of the measures over time (for example, obtained GED within one year) • Measures combining different educational milestones (such as training completion or degree attainment) if individual measures are available • Measures of credit attainment • Measures of educational attainment at a specific institution or group of institutions, unless the group of institutions for which data is available is (1) similar to the set of institutions covered in an established data source (for example, the

Rules for selecting findings based on outcome measures	Notes
	National Student Clearinghouse), or (2) similar to the set of institutions attended by all study participants.[a]

[a] Similarity is judged using the attrition threshold. That is, the authors should demonstrate that the overall and differential differences in the rate at which study participants enroll in the group of institutions should be below the thresholds for overall and differential attrition, as described in Section 3.2.4.

Note: If a study does not examine any listed outcomes within a domain but does examine one or more closely related outcomes, review team leaders will use discretion in selecting findings for review.

Nevertheless, if the Pathways to Work Evidence Clearinghouse reviews a study including outcomes without face validity, those outcomes will receive a low study quality rating.[11] The Pathways to Work Evidence Clearinghouse tracks any outcomes determined to have insufficient face validity in a centralized list so that these outcomes can be systematically excluded across reviews and reviewers. In the event the scope of the Pathways to Work Evidence Clearinghouse expands to include other outcomes for which face validity might be less straightforward to assess, the team will expand these requirements as needed.

3.2.3 Determination of Study Design

The Pathways to Work Evidence Clearinghouse currently reviews research using two eligible designs: RCTs and comparison-group QEDs.

In RCTs, researchers randomly assign study participants to an intervention group or a comparison group.[12] RCTs are considered to produce the strongest evidence possible on effectiveness because random assignment ensures no systematic differences exist between the study groups.

Comparison-group QEDs also use intervention and comparison groups but people are not randomly assigned to them. Instead, researchers typically identify an intervention group that received the program or policy being tested and construct a comparison group that did not receive the services but is otherwise as similar as possible to the intervention group, based on both groups' observed characteristics. Non-random comparison group designs are considered weaker than RCTs because many factors could have led members of the intervention group to choose to receive services and members of the comparison group to choose not to do so. These factors might also have led to differences in outcomes.

As the first step of the review process, reviewers confirm the study used an eligible design and classify the design as an RCT or comparison-group QED. They then proceed to use the appropriate standards (described in the next two sections) to review the study.

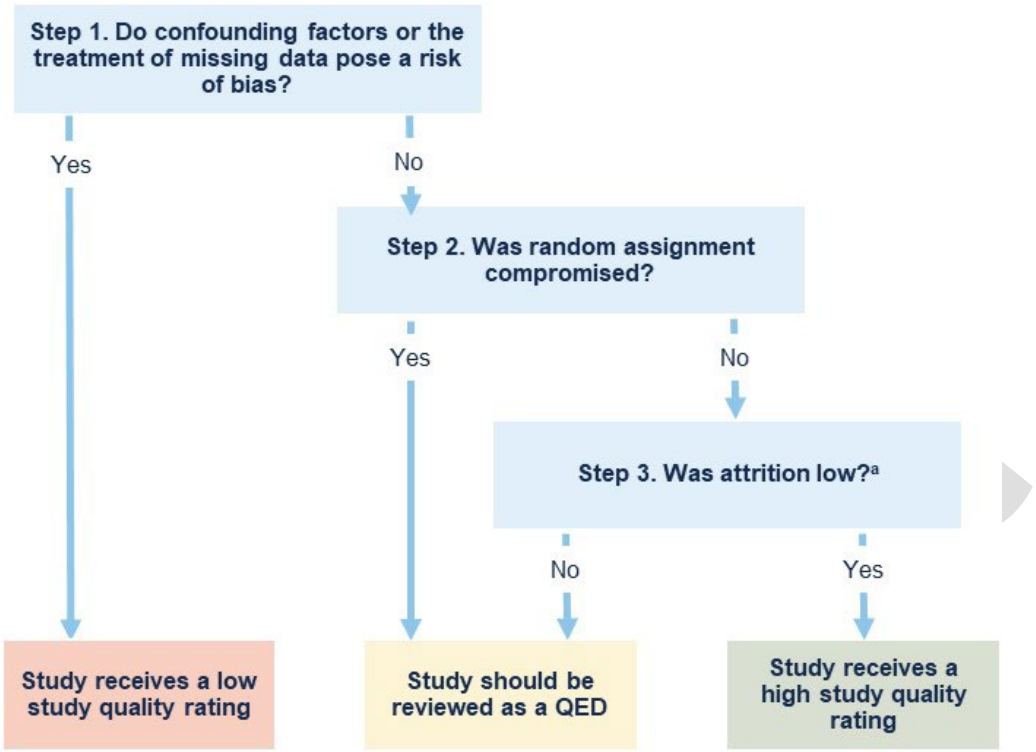
[11] In these rare instances, review team leadership will consult with OPRE to confirm the outcome is not valid.

[12] In Section 3.2, we use the term "intervention" instead of "program" for two reasons. First, the "intervention group" and "comparison group" are commonly used in research design literature to describe how study participants are assigned to conditions. Second, the Pathways to Work Evidence Clearinghouse may review studies where two programs are directly compared to each other. Using the terms "intervention group" and "comparison group", where the intervention group is the program of interest for the review and the comparison group is the condition being contrasted with the program of interest, helps clarify this distinction.

3.2.4 Standards for Reviewing RCTs

Reviewers assess the strength of the evidence provided by RCTs using three main steps, summarized in Exhibit 3.3.

Exhibit 3.3. Process for reviewing RCTs



[a] Based on both differential and overall attrition.

Step 1. Do confounding factors or the treatment of missing data pose a risk of bias?

Two factors can lead an RCT to automatically receive a low study quality rating: confounding factors and mishandling of missing data. **Confounding factors** cause differences between the intervention and comparison groups that cannot be disentangled from the effect of an intervention. One type of confounding factor is an element external to the program that reaches only the members of one study group, for example, if all members of the intervention group lived in one TANF administrative region and all members of the comparison group lived in another TANF administrative region. In this example, it would be impossible to disentangle the effect of the program or policy from that of local economic conditions or local policies. If a confounding factor that is perfectly aligned with one study group is present, a study receives a low study quality rating.

Study authors must also handle missing data appropriately. The most common and straightforward method researchers use when data are missing is to simply remove observations with missing data from the sample they analyze. This approach is called a **complete-case analysis**. But other methods for assessing missing data are sometimes used, including **imputation** (replacing observations with guesses as to the most reasonable value) or **maximum**

likelihood (creating a statistical model to account for the missing data). The What Works Clearinghouse (WWC) Procedures and Standards Handbook Version 5.0 includes all three of these methods as acceptable approaches to handle missing data (WWC, 2022) and also permits use of nonresponse weights, or, for missing regression controls only, replacing the missing data with a constant value and including a missing data indicator in the regression. No additional requirements are placed on the use of any of these five methods. For example, the Pathways to Work Evidence Clearinghouse does not require that nonresponse weights are constructed in a specific manner. Pathways to Work Evidence Clearinghouse reviewers assume a study using any of these five methods handled missing data appropriately. If a study uses a method other than one of these five methods, the Pathways to Work Evidence Clearinghouse principal investigator or review team lead examines the description of the method and determines whether the information is sufficient to ensure that the handling of missing data will not result in biased estimates of intervention effects.[13] If results might be biased due to the handling of missing data, a study receives a low study quality rating.[14]

Step 2. Was random assignment compromised?

A random assignment design is the strongest possible design because the intervention and comparison groups are formed by chance and thus researchers can be confident that observed differences in outcomes are due to the program or policy being studied. If a reviewer identifies deviations from random assignment, such as reassigning or replacing group members after randomization or variations in the probability of random assignment to each condition over time without corresponding adjustment for this variation in the analysis, the Pathways to Work Evidence Clearinghouse treats a study as a comparison-group QED. Otherwise, the review proceeds to Step 3.

Step 3. Was attrition low?

Attrition is the main determinant of whether estimates from an RCT are free of bias and therefore whether the evidence of the program's effectiveness is strong. Attrition refers to the loss of individuals from the study sample over time, or the proportion of the randomly assigned sample not included in the estimation of effects. Both **overall attrition** (percentage of missing cases) and **differential attrition** (how the percentage missing cases differs across the intervention and comparison groups) are a concern, because both might lead to bias in the estimated effects. To determine whether an RCT had low attrition, the project team uses the WWC's *cautious attrition* boundary, defined through an empirical bias model developed for the WWC, based on the levels of overall and differential attrition (see Appendix C).

If an RCT has low attrition, random assignment was not compromised, and no issues are posed by confounding factors or the treatment of missing data, a study receives a high study quality rating. If attrition is high or random assignment was compromised, but there are no issues related

---

[13] The Pathways to Work Evidence Clearinghouse review team lead will centrally document any exceptions to support consistency among review decisions across all research included in the review.

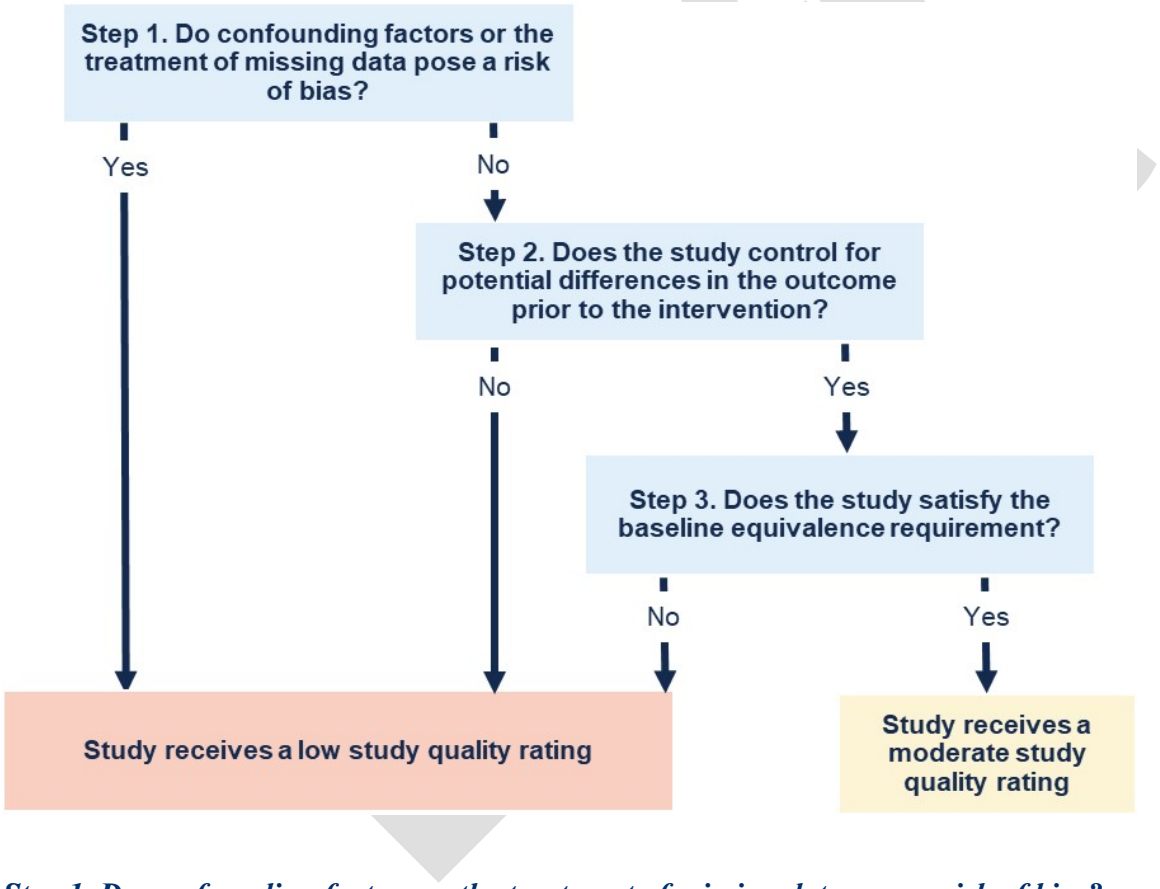
[14] For example, the Pathways to Work Evidence Clearinghouse has determined that hot deck imputation, a common method of imputation not listed by the WWC as acceptable, will produce biased estimates unless combined with a regression approach (see Andridge and Little 2010 for a discussion of the method). Therefore, findings produced using this method receive a low rating.

to confounding factors or missing data, a reviewer proceeds by reviewing the study as if it used a comparison-group QED.

3.2.5 Standards for Reviewing Comparison-Group QEDs

Reviewers assess the strength of the evidence provided by comparison-group QEDs using three steps, summarized in Exhibit 3.4.[15] The highest study quality rating a QED can receive is moderate. Reviewers also use this process to determine whether RCTs that cannot receive a high study quality rating can instead receive a moderate rating.

Exhibit 3.4. Process for reviewing comparison-group QEDs



Step 1. Do confounding factors or the treatment of missing data pose a risk of bias?

This step proceeds in the same manner as Step 1 for reviewing RCTs. See Section 3.2.4 of this chapter for details.

---

[15] Although the Pathways to Work Evidence Clearinghouse's review standards for RCTs closely match those used by CLEAR (2022), which also reviews research on programs and services to improve employment and earnings for people with low incomes, standards differ somewhat for comparison-group QEDs between these two reviews. CLEAR uses a set of more general regression analysis criteria to review comparison-group QEDs and relaxes some of the criteria that the Pathways to Work Evidence Clearinghouse applies in cases where groups are similar at baseline or authors compare intervention and comparison groups over multiple periods. In addition, CLEAR requires reviewers to assess whether sample members' anticipation of an intervention could bias results. The Pathways to Work Evidence Clearinghouse does not use this criterion.

Step 2. Does the study control for any potential differences in the outcome before the intervention?

To receive a moderate rating, a study must control for potential differences in the outcome before the program began (the period just prior to program beginning for the intervention group is often referred to as baseline). Studies typically control for pre-intervention characteristics by including control variables in their regression analysis.[16] If there is no variation in the outcome before the intervention between the intervention and comparison condition, a control is not required in the model (e.g., if the study demonstrates that both the intervention and comparison condition had zero earnings over the relevant eligible period before the program began).

If the outcome examined is an earnings or employment measure, this control should be measured at least one year before the program began. Observations from one year or more before baseline are required because people commonly experience a dip in earnings before program enrollment.[17] For all other outcomes, including benefit receipt and education or training outcomes, this control should be measured shortly before the intervention or study began.

An exception to the requirement for controls for earnings and employment outcomes to be measured at least one year before the program began is made for studies where a dip in earnings in the year prior to the program beginning is highly unlikely. These studies must still control for potential differences in the outcome before the program began – the exception only applies to the timing of when the control was measured. Currently, there are two cases that are exempt from the requirement for the timing of the control measurement: (1) studies where participants have been incarcerated for one year or longer prior to the program beginning; and (2) studies of minor youth who would not have had sufficient time to establish one year or more of work history (e.g., a program designed for 16-year-olds with no prior work history). The requirement for the control to be measured at least one year before the program began may also potentially be waived in cases where authors provide explicit credible evidence that participants in both the intervention and comparison conditions in the study could not have had a change in earnings in the year prior to the program beginning. Waiving this requirement requires the approval of the principal investigator.

If a comparison-group QED (or an RCT that cannot receive a high study quality rating) does not control for the pre-program measure, reviewers assign it a low study quality rating. Otherwise, reviewers proceed to the final step of the review process.

Step 3. Does the study satisfy the baseline equivalence requirement?

To produce credible evidence, comparison-group QEDs must demonstrate that the intervention and comparison groups had similar characteristics at baseline (that is, before intervention group members received program services). Ensuring that the groups were similar before one group was able to receive services helps establish that differences observed between the two groups

---

[16] Pathways to Work follows the current WWC procedures and standards to determine which methods of adjustment for potential differences in pre-intervention outcomes are acceptable. Controlling for a propensity score summarizing the probability of group assignment (rather than directly controlling for the baseline or lagged measures used to construct the propensity score) is not an acceptable method of controlling for pre-intervention outcomes.

[17] First formally documented in Ashenfelter (1978), and commonly referred to as the preprogram dip or Ashenfelter dip. See Heckman and Smith (1999) for details.

after receiving services were actually the result of the program examined. The Pathways to Work Evidence Clearinghouse requires studies to demonstrate baseline equivalence based on five characteristics: [18]

1) earnings (or employment for employment outcomes), measured one year or more before baseline (unless the criteria for an exemption from the timing requirement are met, as described above);
2) a measure of socioeconomic status (such as educational attainment or receipt of some means-tested public benefit, such as food stamps), measured shortly before program services or study began;[19]
3) race and ethnicity;
4) gender; and
5) age.

The Pathways to Work Evidence Clearinghouse first assesses equivalence based on whether the difference in means for the characteristic across the two groups is not statistically significant at the 0.05 level (using a *chi*-squared test for categorical variables and a two-tailed *t*-test otherwise). If the test is not statistically different, the baseline equivalence requirement is met.

If the difference is statistically significant, reviewers then compute the effect size of the baseline difference, following the procedures specified in Appendix D. If the effect size is less than 0.05 standard deviations, the baseline equivalence requirement is met. If the effect size is between 0.05 and 0.25 standard deviations, the study must control for the characteristic in their statistical model. If the effect size is greater than 0.25 standard deviations (or if an effect size cannot be computed), then the baseline equivalence requirement is not met.

Comparison-group QEDs, and RCTs with high attrition or compromised random assignment, that have at least one outcome that meets the control and baseline equivalence requirements, and do not have any issues related to confounding factors or missing data, receive a moderate study quality rating. Otherwise, reviewers assign these studies a low study quality rating.

3.3. The Review Process

The Pathways to Work Evidence Clearinghouse team implements the following processes to support consistency and quality of reviews.

---

[18] To avoid overburdening study authors, Pathways to Work Evidence Clearinghouse reviewers may assess baseline equivalence using information for a sample of individuals that differs slightly from the sample of individuals used to produce a finding (for example, due to item-level nonresponse on a survey), so long as the difference in samples falls below the threshold for high attrition (see Section 3.2.4, Step 3 of this chapter).

[19] If multiple measures of socioeconomic status are available to use in assessing baseline equivalence, reviewers examine the broadest measure possible (for example, choosing receipt of any public benefits over receipt of TANF benefits). If no measure is clearly preferred, the following priority order should be used: years of education, educational attainment in categories, share that did not attain a high school diploma or equivalent certificate, share receiving any public benefits, share receiving SNAP or Food Stamp benefits, share receiving TANF or other cash assistance, and share receiving Medicaid or other means-tested, publicly-funded medical services.

3.3.1 Study Review Process

The study review process was designed to ensure every study is carefully considered and assigned the most appropriate study quality rating. Two Pathways to Work Evidence Clearinghouse reviewers examine each study selected for review. The first reviewer documents all relevant information and assigns a preliminary study quality rating. The second reviewer thoroughly checks the review to make sure the study quality rating criteria were correctly applied, and the review captured all appropriate information. When the first and second reviewers are not certain of a rating or come to different conclusions, they further consult with the review team lead.

Some studies do not contain all the information desired for the review effort. When key information is missing, the review team requests it from study authors.[20] If study authors do not provide the requested information, the review team makes the most conservative assumptions that the information provided can support. For example, if the information needed to assess attrition is not available, the team will assume attrition is high. Reviewers also document whether a study might have received a higher rating if additional information had been available.

3.3.2 Challenges to Review Findings

The Pathways to Work Evidence Clearinghouse Quality Review Team (QRT) handles any challenges interested parties make about a review's findings, the inclusion of a study within the Pathways to Work Evidence Clearinghouse, or other individual judgements the Pathways to Work Evidence Clearinghouse team makes. The QRT addresses any issues with reviews that these parties raise, so long as they are (1) submitted in writing to PathwaysClearinghouse@abtglobal.com, (2) related to a specific study or well-defined set of studies, and (3) coherently explained (and the inquirer is available to answer any clarifying questions).

When a request is submitted to the QRT, a team member first verifies the request meets the criteria listed above. After this confirmation, the team member examines the study and any related materials, discusses the review with the original study reviewers, and presents a summary of the review and any potential flaws to the QRT. The QRT then determines whether the initial review should be revised, notifies OPRE and the inquirer of its findings and, if necessary, edits any Clearinghouse products to reflect the updated review.[21]

3.3.3 Handling Potential Conflicts of Interest

The Pathways to Work Evidence Clearinghouse requires all reviewers to sign a certification disclosing any potential conflicts of interest, such as having been a member of the research team of a study that the Pathways to Work Evidence Clearinghouse potentially reviews. Reviewers are not permitted to make final decisions related to the results of study reviews or program effectiveness ratings if they have a conflict of interest. If members of the QRT are asked to

---

[20] In some cases, the review team may use information from analyses that authors mention conducting in the manuscript, but do not report the findings from, to determine study ratings or calculate effect sizes. In these cases, the review team may request information about methods used and numerical results from such analyses. The Pathways to Work Evidence Clearinghouse does not ask authors to conduct new analyses.

[21] OPRE will review QRT findings and provide input as needed. However, QRT evaluations of studies funded by HHS (of which OPRE is a part) will be conducted independently by the Pathways to Work Evidence Clearinghouse team and will not be subject to change based on comments from OPRE.

evaluate a study for which they have a conflict of interest, they will consult with other Clearinghouse staff who do not have such conflicts in responding to the QRT request.

3.4. Extracting Data From a Study

Throughout the process of determining study ratings, the Pathways to Work Evidence Clearinghouse team records several key pieces of information at the study, manuscript, and findings levels (see Exhibit 3.5 below). Reviewers use a template to systematically record this data. These templates are then combined to form the Pathways to Work Evidence Clearinghouse database.

Using the findings-level information, the team estimates an **effect size** for each finding with a high or moderate rating and for which sufficient information is provided to do so. Effect sizes are a valuable tool for conducting meta-analyses and syntheses. More broadly, they provide a way to combine or compare results measured in different units across outcomes, programs, and studies. Effect sizes also allow researchers to compare an estimated impact to any statistical distribution. For example, one could compare the estimated impact of a training program to the overall distribution of earnings for workers in the program's target population in a specific year. That is, one can use effect sizes to estimate an average effect across multiple studies and outcomes, and then transform that average back into an easy-to-understand number, such as dollars or percentage points. The Pathways to Work Evidence Clearinghouse takes this approach to summarize results.

The team calculates effect sizes as **Hedges' *g***, the ratio between the estimated impact of the intervention and the standard deviation pooled across intervention and comparison groups.[22] To avoid bias due to small sample sizes, team members also apply a sample size correction to effect sizes (the correction gets vanishingly small as the sample size grows).

Because, for many findings, the standard deviations needed to calculate effect sizes are not available, Pathways to Work Evidence Clearinghouse reviewers attempt to calculate two effect sizes for each finding rated high or moderate. These effect sizes include one that uses a study-specific standard deviation to normalize the impact estimate, and one that uses a national standard deviation to normalize the impact estimate (see Box 3.2 below). The team calculates an effect size using study-specific data if the Pathways to Work Evidence Clearinghouse can obtain from the study authors the information needed to do so. The team calculates an effect size using national data when a nationally representative measure of the standard deviation of the outcome can be constructed. Appendix D provides the details on the calculation of these effect sizes.

---

[22] Some systematic reviews use alternative indices to estimate effect sizes for binary variables (that is, those that take on values of 0 and 1 only, such as employment). For example, the WWC uses the Cox index. Although research has shown that using the Cox index can be preferable to using Hedges' *g,* this research is based on assumptions that are unlikely to hold for the key binary outcomes of interest to Pathways to Work (Sánchez-Meca et al., 2003). For example, the Cox index produces artificially large effect sizes when most people in the sample have a 0 or 1 value for the outcome of interest, a condition that is likely to hold for employment and benefit receipt in some studies the Pathways to Work Evidence Clearinghouse reviews (where some studies will have a very high proportion receiving benefits or a very low proportion employed). The Pathways to Work Evidence Clearinghouse therefore uses Hedges' *g* for both binary and nonbinary variables.

Exhibit 3.5. Key data elements collected during study review process

Study-level	• Design • Target populations • Sample characteristics • Setting • Analytic methods • Time period over which the evaluation occurred (enrollment and follow-up periods) • Program history • The program tested and its implementation • Services received by the comparison group • Outcomes examined but not selected for review
Manuscript-level	• Citation • Authors • Year of publication • URL

Findings-level	• Measure
	• Outcome
	• Data source
	• Sample description
	• Sample sizes
	• Baseline means and standard deviations
	• For findings rated high or moderate:
	  ▫ Means and standard deviations of outcome measures
	  ▫ Estimated impacts
	  ▫ Information related to the precision and statistical significance of estimates

Note: To promote efficiency, for studies rated low (see Section 3.2), reviewers only record information at the study-level on the intervention and comparison group conditions (including the services examined, program name, and whether any services were mandatory); limited information on the setting, time period, and target population for the study; the citations for associated manuscripts; and the information leading the study to receive a low rating.

Box 3.2. Comparing the two measures of the effect size

The Pathways to Work Evidence Clearinghouse team attempts to calculate two effect sizes for each finding rated high or moderate: one calculated using a study-specific standard deviation (g_{study}), and one calculated using a standard deviation based on data from the Current Population Survey, a nationally representative survey of U.S. households that has consistently collected information on income and employment since 1962 ($g_{national}$). Using study data to calculate effect sizes produces estimates of the size of a program's effects relative to variation in the outcome for study participants. Using national data to calculate effect sizes produces estimates of the size of effects relative to variation in the outcome across the set of all individuals in the U.S. with low incomes.

To highlight the differences in these measures, consider two studies that both examined a program that increased quarterly earnings by $300. Study 1 includes a reasonably homogenous population with a low standard deviation of earnings, and Study 2 includes a more diverse population with a higher standard deviation of earnings. The two studies would have the same value of $g_{national}$, but the value of g_{study} for Study 1 would be higher than the value of g_{study} for Study 2.

We are not aware of other clearinghouses that have used national data to estimate effect sizes. The Pathways to Work Evidence Clearinghouse team will therefore use the findings for which it can compute both effect sizes to compare the different measures. This will enable the team to determine the circumstances under which the measures produce similar results and when and why results might differ.

Each finding catalogued and determined to have a high or moderate rating is also categorized based on its **sign** (positive or negative), **statistical significance** (statistically significant at the 0.05 level or *null***,** meaning not statistically significant), and **size** (small, or medium or large). The Pathways to Work Evidence Clearinghouse also labels the **direction** of findings as *favorable* or *unfavorable*. Favorable effects are associated with increases in employment, earnings, and education and training or decreases in public benefit receipt. Unfavorable effects are associated with decreases in employment, earnings, and education and training or increases in public benefit receipt.[23] Findings with a low rating are not categorized in this way or otherwise reported.

The Pathways to Work Evidence Clearinghouse classifies an impact as medium or large if its corresponding effect size is more than 0.25 standard deviations. The cutoff should be thought of as providing a reasonably high bar. An effect size of 0.25 would correspond to an increase in earnings of approximately $6,350 in 2023 dollars. Moreover, for a population with a typical employment rate around 80 percent, an effect size of 0.25 would correspond to an increase in employment of about 10 percentage points. For a population with a typical employment rate around 50 percent, an effect size of 0.25 would correspond to an increase in employment of about 12 percentage points. According to Card et al. (2017), the average labor market program for disadvantaged workers raised the employment rate by about 5 percentage points.

[23] The Pathways to Work Evidence Clearinghouse interprets reduced public benefits as a favorable finding because of the overall focus of the project on identifying effective ways to help people with low incomes move from public benefit receipt to employment.

4. Assessing the Evidence of Effectiveness for a Program

The Pathways to Work Evidence Clearinghouse aims to create a repository of programs with rigorous evidence of effectiveness. This chapter describes the approach to synthesizing study evidence in this manner. Section 4.1 provides details on how studies are grouped into programs and Section 4.2 describes how programs are classified based on effectiveness. Section 4.3 provides an overview of the presentation of findings.

4.1. Grouping Studies Into Programs

The *Federal Register* Notice (FRN) associated with the Pathways to Work Evidence Clearinghouse, 83 FR 26290, defines a program as "a specific bundle of services and/or policies implemented in a given context" (p. 26291). Following this, the Pathways to Work Evidence Clearinghouse defines a program based on the services offered to the intervention condition but not offered to the comparison condition.[24] That is, two studies are considered to examine the same program only if the same services were offered in both cases. For the purposes of implementing these criteria, the Pathways to Work team has defined service categories that can be found on the Pathways to Work website.[25] In addition, studies in which participation was mandatory (for continued benefit receipt or other reasons) are classified as examining different interventions from studies with voluntary participation.

In some cases, studies might examine the same services with the same participation requirements but implemented in fundamentally different ways. For example, two programs might provide both training and case management, with one providing one year of training paired with monthly case management meetings and the other providing a two-week training paired with bi-weekly case management meetings for one year. Although the programs provide the same services, the intensity and focus of service delivery varies greatly.

Therefore, the Pathways to Work Evidence Clearinghouse team first groups studies examining the same services and having either mandatory or voluntary participation but then, the principal investigator and review task lead determine whether the studies within a group examine the same program. When studies in a group examine services with fundamentally different theories of change or approaches, the team recommends grouping these studies into multiple programs. These choices are subject to the review and approval of the Pathways to Work Evidence Clearinghouse project director and deputy project director. Clearinghouse staff also consult, as needed, with senior content experts on the Pathways to Work Evidence Clearinghouse team and OPRE about how to sort studies into programs.

In addition to classifying all program services, the Pathways Clearinghouse also designates one service as a program's **primary service**. A program's primary service is the principal service of the program. To identify primary services, reviewers examine each study and identify the service provided as part of the examined program (1) that a large proportion of intervention group members received and a large proportion of comparison group members did not and (2) was described by the study authors as most integral to the theory of change tested by the study. Both

[24] See Chapter 2 for a discussion of the context in which an intervention must have taken place for it to be eligible for review by the Pathways to Work Evidence Clearinghouse.

[25] Although we refer to these elements collectively as services for ease of exposition, a small number might instead be classified as public benefit policies.

the first and second study reviewer independently assess a program's primary service and discuss the study until they achieve consensus.

4.2. Determining Programs' Evidence of Effectiveness

The Pathways to Work Evidence Clearinghouse aggregates information from across studies to determine a program's **effectiveness rating** within each of the outcome domains. Possible ratings, defined in Exhibit 4.1 below, fall into six categories:

1. Programs receiving the **Well-Supported** rating in a domain are those in which the evidence indicates a program is likely to improve outcomes in a domain if the program was replicated. Findings rated high or moderate from at least two studies conducted in the United States must show favorable and statistically significant effects, with no strong countervailing evidence, for this rating to be assigned.[26] However, because implementation challenges and successes often vary, and no two implementations of a program are identical, Pathways to Work Evidence Clearinghouse users should not view this rating as a guarantee of success.
2. Programs receiving the **Supported** rating in a domain are those with more limited evidence of success within the domain. These programs have at least one study with findings rated moderate or high showing evidence of favorable and statistically significant effects in the domain, but the evidence is less conclusive than that for well-supported programs.
3. Programs receiving the **Mixed Support** rating in a domain are those with some evidence from findings rated moderate or high indicating that they improve outcomes, and some evidence from findings rated moderate or high indicating that they worsen outcomes. These programs might produce positive or negative effects, depending on contextual and implementation factors.
4. Programs receiving the **Insufficient Evidence to Assess Support** rating in a domain are those that have a single study where none of the findings rated moderate or high are statistically significant. These programs lack a sufficient body of evidence to receive one of the other ratings and require further study to support conclusions about their effectiveness.
5. Programs receiving the **Not Supported** rating in a domain are those that demonstrate a pattern of null and/or unfavorable effects from findings rated moderate or high. These programs are not likely to improve outcomes if implemented in contexts similar to those used in prior research.
6. Programs receiving the **Cannot Assess Support** rating in a domain are those that have no findings rated moderate or high and therefore cannot receive a rating. These programs also require further study to support conclusions about their effectiveness.

---

[26] Although the Pathways to Work Evidence Clearinghouse includes research conducted in Canada, its chief goal is to inform decision makers working within the policy environment of the United States. Therefore, to receive the highest effectiveness rating, a program must demonstrate evidence of effectiveness in the United States. Programs tested only outside of the United States can still receive the Supported rating.

Exhibit 4.2 below provides further examples of how the team would rate a program within a domain based on all possible combinations of findings from two studies that each included one finding in the domain.

Exhibit 4.1. Potential effectiveness ratings of programs, by domain

Program-domain designation	Requirements
Well-Supported	Based on all studies with a moderate or high study quality rating conducted in the United States, and within findings receiving a high or moderate rating: • There are favorable and statistically significant findings in the domain from two or more studies of the program. AND • There are no unfavorable and statistically significant findings in the domain. AND • Across all findings in the domain, the average effect size (weighted by the sample sizes used to estimate the effect) is favorable. AND • There are at least as many impacts in the domain that are favorable and either statistically significant or medium or large, as the number that are either unfavorable or favorable and small.
Supported	Based on all studies with a moderate or high study quality rating, and within findings receiving a high or moderate rating: • The program has at least one favorable and statistically significant finding in the domain. AND • The program has no unfavorable and statistically significant findings in the domain. AND • The program does not meet the criteria for being classified as Well-Supported within the domain.
Mixed Support	Based on all studies with a moderate or high study quality rating, and within findings receiving a high or moderate rating: • The program has at least one favorable and statistically significant finding in the domain. AND • The program has at least one unfavorable and statistically significant finding in the domain.
Insufficient Evidence to Assess Support	Based on all studies with a moderate or high study quality rating and within findings receiving a high or moderate rating: • There is a single study with a moderate or high study quality rating examining findings in the domain. AND • The study has only null findings in the domain.
Not Supported	At least one study has been conducted and received a moderate or high study quality rating and examined findings in the domain that received a high or moderate rating: AND None of the above effectiveness ratings apply.
Cannot Assess Support	There is no study that included outcomes in the domain that received a moderate or high study quality rating.

Note: If authors provide separate estimates for subgroups of people defined by any characteristic other than site or timing of service receipt, the subgroup estimates will be treated as if they are from the same study for the purposes of assigning program effectiveness ratings (but will be catalogued as separate studies on the Pathways to Work Evidence Clearinghouse website). Studies that focus on a specific site or cohort of individuals are treated as separate studies for all purposes.

Exhibit 4.2. Potential designations of a program with two studies

Study A: Designation of single finding in domain	Study B: Designation of single finding in domain	Program rating
Favorable, statistically significant	Favorable, statistically significant	Well-Supported
Favorable, statistically significant	Favorable, null, medium or large	Supported
Favorable, statistically significant	Null, small	Supported
Favorable, statistically significant	Unfavorable, null, medium or large	Supported
Favorable, statistically significant	*Did not examine outcomes in domain*	Supported
Favorable, statistically significant	Unfavorable, statistically significant	Mixed Support
Favorable, null, medium or large	*Did not examine outcomes in domain*	Insufficient Evidence
Favorable, null, small	*Did not examine outcomes in domain*	Insufficient Evidence
Unfavorable, null, medium or large	*Did not examine outcomes in domain*	Insufficient Evidence
Favorable, null, medium or large	Favorable, null, medium or large	Not Supported
Favorable, null, medium or large	Null, small	Not Supported
Favorable, null, medium or large	Unfavorable, null, medium or large	Not Supported
Favorable, null, medium or large	Unfavorable, statistically significant	Not Supported
Null, small	Null, small	Not Supported
Null, small	Unfavorable, null, medium or large	Not Supported
Null, small	Unfavorable, statistically significant	Not Supported
Unfavorable, null, medium or large	Unfavorable, null, medium or large	Not Supported
Unfavorable, null, medium or large	Unfavorable, statistically significant	Not Supported
Unfavorable, statistically significant	Unfavorable, statistically significant	Not Supported
Unfavorable, statistically significant	*Did not examine outcomes in domain*	Not Supported
Did not examine outcomes in domain	*Did not examine outcomes in domain*	Cannot Assess

Note: If authors provide separate estimates for subgroups of people defined by any characteristic other than site or timing of service receipt, the subgroup estimates will be treated as if they are from the same study for the purposes of assigning intervention effectiveness ratings (but will be catalogued as separate studies on the Pathways to Work Evidence Clearinghouse website). Studies that focus on a specific site or cohort of individuals are treated as separate studies for all purposes.

As in any systematic review, there is some risk that statistical bias could lead the Pathways to Work Evidence Clearinghouse to come to incorrect conclusions about intervention effectiveness. However, two key elements of the approach limit this risk of bias. First, only studies and findings deemed to provide moderate- or high-quality evidence are used to assign program ratings (other than those receiving Cannot Assess Support due to a lack of moderate or high quality evidence). Studies rated as high and moderate are those in which the extent of bias is unlikely to be sufficiently large to alter the studies' main conclusions. Second, only interventions for which multiple studies reach the same conclusions about intervention effectiveness can receive the rating of Well-Supported. Consistent with the recommendations of the Cochrane Collaboration, we do not attempt to further summarize the overall risk of bias for the review effort as a whole (Higgins and Green, 2011).[27]

4.3. Presenting Findings

The Pathways to Work Evidence Clearinghouse website is powered by a searchable database with elements at the program, study, manuscript, and finding levels. Each program also has a dedicated web page, clearly indicating the findings that the Pathways to Work Evidence Clearinghouse reviewed from each of the outcome domains. Users can learn more about the studies of each program, and about specific findings, on the program page and by navigating to additional pages.

The Pathways to Work Evidence Clearinghouse team will continue developing synthesis products designed to further explain what services and policies work for whom, and under what conditions. Topics will be selected based on input from ACF and from Pathways to Work Evidence Clearinghouse audiences and expert panels. These might focus on different approaches, or guiding frameworks, for providing services. Syntheses might also focus on common themes, such as specific populations, barriers to employment, or local conditions. Methods for each synthesis will be specified before work on that synthesis begins, in accordance with the PRISMA-P guidelines (Moher et al., 2015).

[27] The Cochrane Collaboration is an international not-for-profit organization that produces systematic reviews to help people make informed health decisions. The Cochrane Collaboration Handbook is a leading resource on the conduct of systematic reviews.

5. Assessing Cost Information For A Program

As the evidence base on employment programs for job seekers with low incomes continues to grow, so has the need for information about the costs of those programs and practices. Without additional information about the personnel/non-personnel resources used and the associated costs, it is impossible to provide guidance on the resources necessary for implementation and how best to allocate funding towards these efforts. The information required to provide this guidance is best generated by rigorous analysis of program cost, cost-effectiveness, and/or cost-benefit analyses. This chapter contains standards for reviewing cost studies on employment and training programs designed for individuals with low incomes and included in the Pathways to Work Evidence Clearinghouse.

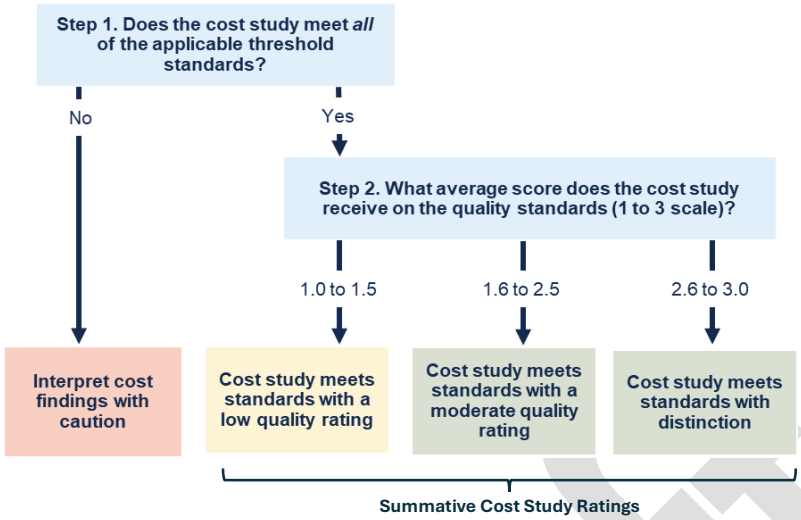
Cost studies can include cost analysis, cost-effectiveness analysis, and cost-benefit analysis. **Cost analysis** provides an analysis of the comprehensive effort involved in program implementation, answering questions such as how much the program costs to implement, the per-participant cost of the program, and the feasibility of implementation given existing budget constraints and available resource inputs. **Cost-effectiveness analysis** compares the estimated cost of a program with an estimate of its impact on a given outcome of interest. **Cost-benefit analysis** compares the cost of a program with the monetized outcomes associated with that program.[28]

Section 5.1 describes *threshold standards* that identify basic characteristics that cost studies must meet for the Pathways to Work Evidence Clearinghouse to rate the quality of the study's cost analysis information. A cost study that meets the threshold standards for a cost analysis indicates the study possesses the fundamental characteristics necessary to produce accurate cost information. Section 5.2 describes *quality rating standards* that are applied to studies that meet the threshold standards and provide additional information on the quality of the cost analysis estimates. All cost studies will be reviewed using the applicable standards for the type(s) of research conducted. Studies with cost analyses are reviewed using the standards in Section 5.1 and 5.2 below. The standards used to rate cost-effectiveness and cost-benefit analyses are provided in Appendix E.[29] Exhibit 5.1 below summarizes how cost studies will be evaluated using the threshold and quality standards to produce a summative cost study rating. As an example, if a cost study includes a cost analysis and a cost-effectiveness analysis, it would be reviewed using the cost analysis standards in this section and receive a summative rating based on the cost analysis, and it also would be reviewed using the cost-effectiveness standards in Appendix E and receive a separate summative rating for the cost-effectiveness analysis. Appendix F provides a glossary of terms for cost studies and the standards in this chapter. Glossary terms in this chapter are bolded and underlined for easy reference.

[28] For more information about cost analysis, cost-effectiveness, and cost-benefit analysis, please see the *Standards for the Economic Evaluation of Educational and Social Programs* (CASP, 2021).

[29] The cost-effectiveness and cost-benefit threshold and quality rating standards largely overlap with the cost analysis standards, but there are some additional standards that apply only to these types of analyses. The same scoring process is applied to produce separate summative ratings for the cost-effectiveness and/or cost-benefit analyses presented in the cost study.

Exhibit 5.1. Process for reviewing cost studies



Note. Average scores for summative cost study ratings are rounded to the nearest tenth prior to assigning the rating.

5.1. Threshold Standards for Cost Analyses

Reviewers first apply a set of threshold standards to determine whether a cost analysis can be rated for its quality. For each threshold *standard*, reviewers choose "yes" or "no" based on whether a cost analysis meets the *criteria* defined for the standard. The standards are numbered, with each standard having a list of criteria (e.g., a, b, c) that must be satisfied for a standard to be met.

1. The cost analysis study describes the program of interest by satisfying the following two criteria:
    a. The study includes a description of the key program components as well as the context in which the program was implemented (e.g., specific geographic location of the program sites).
    b. The study includes a description of the program participants, including the number of participants and any available background characteristics (e.g., participant age, education level, socioeconomic status, income level, race/ethnicity, gender), preferably by site.
2. The cost analysis study indicates that program costs are estimated using a **resource-based approach** that accounts for all of the **personnel resources** and **non-personnel resources** used for implementation. The following three criteria must be satisfied:
    a. The study demonstrates that a valid resource-based approach was used to estimate implementation costs (e.g., application of the **ingredients method**, a resource cost model).
    b. The study accounts for the full set of resources used to implement the program, including any resources that were purchased, redirected to the program from another purpose, or donated/volunteered.

c. The study provides a description of the personnel and non-personnel resources accounted for in the cost estimate, including the quantities of the resources used and their associated prices.

3. The cost analysis study reports for which entities costs are being calculated (i.e., the analysis **perspective**).

    a. The study clearly states the perspective(s) from which costs are represented. This could be the societal perspective, the public perspective, the private perspective, or the perspective of more granular groups, such as specific agencies, funders, or program participants.

4. The cost analysis study estimates and reports the per-participant cost of the program, including the site-specific per-participant costs and/or the average cost per participant across sites.

    a. The study includes at least one of the following: (1) the site-specific per-participant cost for each site included in the cost study, or (2) the overall average per-participant cost of the program, including how the site-specific per-participant costs were used to estimate the average per-participant cost across sites (e.g., weighting by the number of participants served versus taking a straight average to estimate costs across sites).

5.2. Quality Rating Standards for Cost Analyses

Cost analysis studies that meet all applicable threshold standards will then be evaluated by reviewers using the cost analysis quality rating standards, which provide additional information on the quality of the cost estimates. Each quality rating standard is scored on a scale from 1 to 3, where 3 is the highest quality rating (some quality standards only have two possible rating scores of 1 or 3). At the end of the review, a summative cost analysis quality rating score is determined based on the average quality rating scores across all quality standards.[30] The summative rating score is then rounded to the nearest tenth and converted into one of three cost study summative ratings[31], illustrated in Exhibit 5.1 above:

- **Cost Study Meets Standards With Low Quality**. Assigned for an average quality score between 1.0 and 1.5.
- **Cost Study Meets Standards With Moderate Quality**. Assigned for an average quality score between 1.6 and 2.5.
- **Cost Study Meets Standards With Distinction**. Assigned for an average quality score between 2.6 and 3.0.

Cost Analysis Quality Rating Standards

[30] Cost-effectiveness and cost-benefit analyses that meet their respective threshold standards receive separate quality rating scores using the same procedures as cost analyses to assign a summative rating for the cost-effectiveness and/or cost-benefit analysis (see Appendix E).

[31] For example, summative rating scores of both 2.48 and 2.52 would both be rounded to 2.5 and receive a "Moderate Quality" rating; a score of 2.58 would be rounded to 2.6 and receive a "With Distinction" rating.

1. The cost analysis study specifies whether the prices used to estimate costs reflect actual **local prices** or have been standardized to state or national average price levels. Scored as either 1 or 3, as follows:
    - (1) The study does not mention whether prices are local or represent standardized state or national prices.
    - (3) Prices are reported as being local or standardized to represent some other geographic area.
2. The cost analysis study uses prices that represent the **opportunity cost** of the resources used, based on prices found in competitive markets or a justifiable **shadow price**. Scored as:
    - (1) The study provides no information on the prices used to calculate the costs of resources.
    - (2) The study includes the prices that were used to calculate resource costs.
    - (3) The study includes the prices that were used to calculate resource costs and lists the pricing sources.
3. The cost analysis study uses prices that represent a common year for the resources included in the cost estimate. Scored as:
    - (1) The study provides no information on whether the resource prices represent a common year.
    - (2) The study includes information showing that prices used to calculate program costs represent a common year.
    - (3) The study includes information showing that prices used to calculate program costs represent a common year and describes the inflation adjustments used to bring all prices to a common year.
4. The cost analysis study reports the present value of program costs across the analysis period and expresses these for a common reference year. Scored as either 1 or 3, as follows:
    - (1) The study does not report at least one of the following components needed to calculate a present value: (a) the specific year for which costs are represented, (b) the period over which program costs were incurred, and (c) the discount rate used in the calculation of present values.
    - (3) The study reports all of the following components needed to calculate a present value: (a) the specific year for which costs are represented, (b) the period over which program costs were incurred, and (c) the discount rate used in the calculation of present values.
5. The sites included in the cost analysis study are representative of those in which the program of interest is being implemented. Scored as:
    - (1) The study does not include information about site selection for the cost study.
    - (2) The study includes information about the sites included in the cost study but does not explain whether that selection is representative of the sites in which the program was implemented.

- (3) The study provides evidence that the sites included in the cost study are representative of the sites in which the program was studied, or that all sites were included in the cost study. Evidence of representativeness is demonstrated by showing similarities between the analysis program sites and population of sites with respect to any of the following characteristics: background characteristics of participants (e.g., age, education level, socioeconomic status, income level, race/ethnicity, gender), as well as the sites' geographic region and locale (e.g., urban versus rural).

6. The cost analysis study specifies when cost data collection occurred relative to the period when program costs were incurred (i.e., **concurrent**, **retrospective**, or **prospective data collection**). Scored as:
    - (1) The study does not include details about when information to estimate costs was collected.
    - (2) The study indicates that the information used to estimate costs was collected prospectively or retrospectively to when program costs were incurred.
    - (3) The study indicates that information used to estimate costs was collected concurrently with when the program costs were incurred.
7. The cost analysis study reports how program resources are financed or otherwise provided by different entities and how costs are distributed across entities. Examples of entities include program participants, program staff, program funders, or resource providers. In addition to being directly financed, new **resources** may be provided through volunteer time and/or donated non-personnel items, whereas existing resources may be redirected from other programs or purposes. Scored as:
    - (1) The study has no description of how resources were financed or otherwise provided by different entities nor how costs were distributed across different entities.
    - (2) The study either describes how program resources were financed or otherwise provided by different entities, or describes how costs were distributed across entities, but not both.
    - (3) The study describes both how program resources were financed or otherwise provided by different entities, and how costs were distributed across entities.
8. The cost analysis study reports the findings of a **sensitivity analysis** that evaluates how results change in response to alternative assumptions used to estimate costs. Assumptions may relate to types or quantities of resources used and their prices, magnitude of outcome impact, characteristics of the participants served, sites included in the analysis, discount rates used to calculate present values, the time horizons over which costs are calculated, or other relevant assumptions. Scored as:
    - (1) The study has no mention of a cost study sensitivity analysis being conducted.
    - (2) The study mentions that a cost study sensitivity analysis was conducted but does not provide a detailed description of the findings of the analyses.

- (3) The study provides an in-depth description of least one cost study sensitivity analysis that was conducted, including what assumptions were explored and the degree to which the study findings were sensitive to these assumptions.

9. The cost analysis study describes whether key program resources correspond to **fixed costs**, **variable costs**, or **lumpy costs**, for the purpose of describing the cost implications of implementation at different scales. Scored as:
    - (1) The report does not include information about whether the key resources used to implement the program are associated with fixed, variable, or lumpy costs.
    - (2) The study includes information about whether the key resources used are associated with fixed, variable, or lumpy costs but does not use this information to discuss the cost implications of implementation at different scales.
    - (3) The study includes information about key resources used being related to fixed, variable, or lumpy costs and uses this information to discuss the cost implications of implementation at different scales.

References

Andridge, Rebecca R., and Roderick J. Little. "A Review of Hot Deck Imputation for Survey Non-response." *International Statistical Review*, vol. 78, no. 1, 2010, pp. 40–64.

Ashenfelter, Orley. "Estimating the Effect of Training Programs on Earnings." *Review of Economics and Statistics,* vol. 60, 1978, pp. 47–57.

Card, David, Jochen Kluve, and Andrea Weber. "What Works? A Meta-Analysis of Recent Active Labor Market Program Evaluations." *Journal of the European Economic Association*, vol. 16, no. 3, 2017, pp. 894–931.

Clearinghouse for Labor Evaluation and Research (CLEAR). "CLEAR Causal Evidence Guidelines, Version 2.2." Washington, DC: U.S. Department of Labor, 2022. Available at https://clear.dol.gov/sites/default/files/CLEAR%20Causal%20Evidence%20Guidelines_v.2.2_1.pdf. Accessed September 24, 2024.

Crowley, D. Max, Kenneth A. Dodge, W. Steven Barnett, Phaedra Corso, Sarah Duffy, Philip Graham, Mark Greenberg, Ron Haskins, Laura Hill, Damon E. Jones, Lynn A. Karoly, Margaret R. Kuklinski, and Robert Plotnick. "Standards of Evidence for Conducting and Reporting Economic Evaluations in Prevention Science." *Prevention Science*, vol. 19, no. 3, 2018, pp. 366–390.

Colquhoun, Heather L., Danielle Levac, Kelly K. O'Brien, Sharon Straus, Andrea C. Tricco, Laure Perrier, Monika Kastner, and David Moher. "Scoping Reviews: Time for Clarity in Definition, Methods, and Reporting." *Journal of Clinical Epidemiology,* vol. 67, no. 12, 2014, pp. 1291–1294.

Consolidated Appropriations Act of 2017 (Pub. L. 115-31).

Criteria for Evidence of Effectiveness To Be Applied to Projects Identified for Inclusion in the What Works Clearinghouse of Proven and Promising Projects To Move Welfare Recipients Into Work. 83 FR 26290.

Flood, Sarah, Miriam King, Renae Rodgers, Steven Ruggles, and J. Robert Warren. *Integrated Public Use Microdata Series, Current Population Survey: Version 6.0* [dataset]. Minneapolis, MN: IPUMS, 2018.

Guise, J. M., C. Chang, M. Butler, M. Viswanathan, and P. Tugwell. "AHRQ Series on Complex Intervention Systematic Reviews—Paper 1: An Introduction to a Series of Articles that Provide Guidance and Tools for Reviews of Complex Interventions." *Journal of Clinical Epidemiology*, vol. 90, 2017a, pp. 6–10.

Guise, J. M., M. E. Butler, C. Chang, M. Viswanathan, T. Pigott, and P. Tugwell. "AHRQ Series on Complex Intervention Systematic Reviews—Paper 6: PRISMA-CI Extension Statement and Checklist." *Journal of Clinical Epidemiology*, vol. 90, 2017b, pp. 43–50.

Heckman, James J., and Jeffrey A. Smith. "The Pre-Programme Earnings Dip and the Determinants of Participation in a Social Programme. Implications for Simple Programme Evaluation Strategies." *The Economic Journal*, vol. 109, no. 45, 1999, pp. 313–348.

Higgins, Julian P.T., and Sally Green. "Cochrane Handbook for Systematic Reviews of Interventions." *Cochrane Collaboration*, 2011. Available at https://handbook-5-1.cochrane.org/front_page.htm. Accessed May 1, 2019.

Irving, Shelley, and Tracy Loveless. "Dynamics of Economic Well-Being: Participation in Government Programs, 2009–2012: Who Gets Assistance?" 2015. Suitland-Silver Hill, MD: United States Census Bureau.

Levin, Henry, Patrick McEwan, Clive Belfield, A. Brooks Bowden, and Robert Shand. *Economic Evaluation in Education: Cost-Effectiveness and Benefit-Cost Analysis (3rd ed.)*. 2018. Sage Publications.

McGowan, Jessie, Margaret Sampson, Douglas M. Salzwedel, Elise Cogo, Vicki Foerster, and Carol Lefebvre. "PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement." *Journal of Clinical Epidemiology*, vol. 75, 2016, pp. 40–46.

Moher, David, Lesley Stewart, and Paul Shekelle. "All in the Family: Systematic Reviews, Rapid Reviews, Scoping Reviews, Realist Reviews, and More." *Systematic Reviews*, vol. 4, 2015, article 183.

Munn, Zachary, Micah D. J. Peters, Cindy Stern, Catalin Tufanaru, Alexa McArthur, and Edoardo Aromataris. "Systematic Review or Scoping Review? Guidance for Authors When Choosing Between a Systematic or Scoping Review Approach." *BMC Medical Research Methodology*, vol. 18, no. 1, 2018, article 143.

Rotz, Dana, Emily Sama-Miller, and Paul Burkander. "Protocol for the Pathways to Work Evidence Clearinghouse: Methods and Standards." OPRE Report # 2020-44, Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2020.

Sánchez-Meca, Julio, Fulgencio Marín-Martínez, and Salvador Chacón-Moscoso. "Effect-Size Indices for Dichotomized Outcomes in Meta-Analysis." *Psychological Methods*, vol. 8, no. 4, 2003, pp. 448–467.

Thompson, Matthew, Arpita Tiwari, Rongwei Fu, Esther Moe, and David I. Buckley. "A Framework to Facilitate the Use of Systematic Reviews and Meta-Analyses in the Design of Primary Research Studies." Rockville, MD: Agency for Healthcare Research and Quality, 2012.

What Works Clearinghouse. "What Works Clearinghouse Standards Handbook Version 5.0." Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Educational Evaluation and Regional Assistance, 2022. Available at https://ies.ed.gov/ncee/WWC/Docs/referenceresources/Final_WWC-HandbookVer5_0-0-508.pdf

Appendix A: Process for Developing and Revising Review Protocol

The Pathways to Work Evidence Clearinghouse was developed in response to the Consolidated Appropriations Act of 2017 (Pub. L. 115-31), which directs the U.S. Department of Health and Human Services (HHS) to create a What Works Clearinghouse of Proven and Promising Projects to Move Welfare Recipients into Work. This appendix describes the process by which the protocol, methods, and standards for Pathways to Work Evidence Clearinghouse were developed. This work occurred in multiple stages.

Building on the Employment Strategies for Low-Income Adults Evidence Review (ESER)

In the initial development stage, staff on the OPRE-sponsored Employment Strategies for Low-Income Adults Evidence Review (ESER) project convened a working group and developed a Federal Register notice (FRN). A group of experts—including representatives from the U.S. Departments of Labor, Education, and Justice, and several HHS offices and agencies, including the Office of Family Assistance, OPRE, the Office of the Assistant Secretary for Planning and Evaluation, and the Agency for Healthcare Research and Quality—met with ESER staff to discuss key parameters for the effort (see 83 FR 26290). The group met several times in fall 2017 and early winter 2018. Key topics discussed included:

- Criteria for classifying interventions as Well-supported and Supported, as well as other classifications of interventions
- Definitions of key terms, including study, intervention, and approach
- Best practices in conducting systematic reviews
- How the Pathways to Work Evidence Clearinghouse can be consistent with other federal review efforts

The ESER team also conducted extensive background research on established methods for reviewing complex interventions and resources related to consistency and transparency in systematic reviews. As a result of these discussions, OPRE issued an FRN (83 FR 26290), which defined the key parameters of the Pathways to Work Evidence Clearinghouse.

Pathways to Work Evidence Clearinghouse, v1.0 (2018-2023)

Next, OPRE developed and competitively awarded a contract to establish the Pathways to Work Evidence Clearinghouse in 2018. To refine the parameters the FRN established, Pathways to Work Evidence Clearinghouse staff met with an expert group in early spring 2019. Discussions focused on how to operationalize key terms and definitions for the Pathways to Work Evidence Clearinghouse and how the scope of the Pathways to Work Evidence Clearinghouse might be expanded beyond that considered by past OPRE systematic reviews. The expert group again included key federal staff, but was further expanded to include research experts on systematic reviews, employment and training services for people with low incomes, and the application of research to policymaking, as well as practitioners and policymakers. The initial methods and standards report (Rotz et al., 2020) reflects these consultations.

Scope of the Pathways to Work Evidence Clearinghouse, v1.0

ESER provided the starting point for the scope of the Pathways to Work Evidence Clearinghouse, with Pathways to Work including all research catalogued by ESER. To be eligible for review under ESER, research had to meet the following criteria:

1. Quantitatively measured the effectiveness of a program using a study design that compared the outcomes for an intervention group (that could receive the program or was subject to the policy) to a similar comparison group (that was not assigned to receive the program or was not subject to the policy). Eligible designs included randomized controlled trials (RCTs) and comparison-group quasi-experimental designs (QEDs).
2. Published in 1990 through mid-2014 (the year the ESER literature search occurred).
3. Estimated the effect of an employment or training program on outcomes related to employment, earnings, public benefit receipt, education, or training.
4. Examined the effects of the program or policy for adults with low incomes (age 18 and older).
5. Examined a program or policy implemented in the United States or Canada (with programs or policies in Canada having been cited by a consulted literature review).[32]

In addition to research reviewed under ESER, the Pathways to Work Evidence Clearinghouse included any literature that would have met these criteria but which was not available to the ESER research team (either because it was not discovered in the ESER literature search and call for studies, or because it was not yet published).

The Pathways to Work Evidence Clearinghouse also expanded upon ESER by (1) including research on youth and adults ages 16 and older (instead of restricting its scope to research on adults age 18 and older) and (2) relaxing the requirement that research on programs implemented in Canada had to have been cited in a consulted literature review. Within this expanded scope, the Pathways to Work Evidence Clearinghouse then prioritized reviewing research previously assessed by ESER and the most recent research within the expanded scope. The Pathways to Work Evidence Clearinghouse website was launched using a database containing all research reviewed by ESER, research published since 1990 related to studies reviewed by ESER, and any eligible studies with initial publication dates in 2010 through 2018.

Pathways to Work Evidence Clearinghouse, v2.0 (2023 – present)

In 2023, OPRE developed and competitively awarded a new contract for the Pathways to Work Evidence Clearinghouse that included revising the Methods and Standards report protocol (here, Protocol Version 2.0) and creating new standards for reviewing information about program costs. A summary of the process used to update the protocol is provided, followed by a summary of updates to the protocol.

Process Used to Update the Methods and Standards Report Protocol

[32] Under ESER, research on interventions conducted in the United Kingdom was also eligible for review if that research was cited in a consulted literature review. However, no such studies were identified.

The Pathways to Work Evidence Clearinghouse team conducted two virtual Evidence Review Expert Consultation (EREC) meetings in summer 2024 to discuss possible revisions and updates to the Pathways to Work Evidence Clearinghouse's review protocol. The project team worked with OPRE to identify and prioritize potential areas for revision and expert consultation in the Methods and Standards Protocol. Key areas prioritized included the literature search, baseline equivalence, program rating categories, and service tag categories and definitions. The panel included methodology experts, subject matter experts, and practitioners with relevant expertise – particularly with respect to the implementation and cost reporting topic areas. Panel members were identified to provide diverse professional experience and perspectives in each of these key priority areas, and were identified through a combination of input from internal experts on Pathways to Work Evidence Clearinghouse team, input from a standing group of field-based practitioners ("Evidence Champions"), and input from OPRE. The purpose of the consultations was to ensure that diverse perspectives and expertise informed the revisions and updates.

The Pathways to Work Evidence Clearinghouse team also worked with OPRE to identify an initial set of elements to be included in the cost standards and met with a separate Cost Expert Advisory Panel (CEAP) in May 2024 to review and discuss these standards. The CEAP included experts in cost analysis and provided input the prioritization of specific cost standards, how draft cost standards needed to be adapted for the Pathways to Work Evidence Clearinghouse, and how study quality ratings could be developed using cost standards. The Pathways to Work team used the EREC and CEAP consultations to further refine these cost standards and discuss options for reporting cost information to the public.

A Federal Register Notice is being issued to invite public comment on the revised protocols and new cost standards. Any additional updates made based on public comments will be summarized in this section in the final version of the protocol.

The scope of the Pathways to Work Evidence Clearinghouse will be periodically revisited. If the scope is changed, the Methods and Standards Protocol will be updated accordingly to ensure that relevant studies and applicable study quality standards are aligned with the updated scope.

Summary of Revisions for Protocol Version 2.0

Introduction (Chapter 1). Some key terms were updated based on active engagement activities, Evidence Champion input, user testing activities, and plain language focus groups. In particular, the term "intervention" was changed to "program" throughout (except in reference to study design, where the use of "intervention" and "comparison" groups is retained).

Literature Search (Chapter 2). Informed by the consultation, the electronic database search protocol was modified (Section 2.1.3), though the study eligibility criteria and scope remained the same. Key changes to the search protocol included: (1) removing geographic terms to better ensure that county- and local-level evaluations are included; (2) removing intent terms, based on findings that search results were not sensitive to these terms; (3) expanding sample terms, with a particular focus on ensuring sufficient breadth of terms for disadvantaged groups; and (4) removing ProQuest Dissertations and Theses database, based on a finding that very few eligible studies were being identified from this database and could be located through alternative literature search strategies (Section 2.1). Minor updates were made to the design and sample terms to ensure comprehensiveness. The list of organizational websites (Section 2.1.4) was updated, and the Prevention Services Clearinghouse was added to the list of other federal evidence reviews used to identify studies (Section 2.1.5).

Assessing A Study's Strength of Evidence (Chapter 3). Two key revisions were made to the standards for reviewing comparison group QEDs (Section 3.2.5). First, the revised standards allow a requirement for earnings to be measured at least one year prior to the start of the intervention to be waived under certain conditions. The revised standards allow for an exemption on this timing requirement in two specific cases: (1) studies where participants have been incarcerated for one year or longer prior to the program beginning; and (2) studies of minor youth who would not have had sufficient time to establish one year or more of work history (e.g., a program designed for 16-year-olds with no prior work history). The timing requirement for the control to be measured at least one year before the program began may also potentially be waived in cases where authors provide explicit credible evidence that participants in both the intervention and comparison conditions in the study could not have had a change in earnings in the year prior to the program beginning. In all cases, this exemption applies only to the timing requirement for when potential differences were measured – not to the requirement to control for any potential differences in the outcome before the program began.

Second, baseline equivalence can now be demonstrated based on the baseline effect size if the baseline difference is statistically significant. If the effect size is less than 0.05 standard deviations, the baseline equivalence requirement is met. If the effect size is between 0.05 and 0.25 standard deviations, the study must control for the characteristic in their statistical model. If the effect size is greater than 0.25 standard deviations (or if an effect size cannot be computed), then the baseline equivalence requirement is not met. This revision is intended to address studies with large sample sizes where even very small differences in the magnitude of the baseline difference may be statistically significant.

Assessing the Evidence of Effectiveness for a Program (Chapter 4). Program rating categories were updated, including updates to the criteria for "insufficient evidence to assess support" and "not supported" categories (Section 4.2).

Assessing Cost Study Information for a Program (Chapter 5). New standards for including and rating the quality of program cost studies are now included, covering cost analysis, cost-effectiveness analysis, and cost-benefit studies. The cost standards include two types of standards: (1) *threshold standards* that identify basic characteristics that cost studies must meet in order to rate the quality of the study's cost information, and (2) *quality rating standards* that are applied to studies that meet the threshold standards and provide additional information about how the study calculated program costs. Cost studies that meet all threshold standards are scored on quality rating standards on a 1 to 3 scale. A summative quality score is then generated by averaging the quality rating scores from each individual quality standard, which also ranges from 1 to 3, and rounding to the nearest tenth. A rating of *Cost study meets standards with low quality* is assigned when the average score is between 1.0 and 1.5. A rating of *Cost study meets standards with moderate quality* is assigned when the average score is between 1.6 and 2.5. A rating of *Cost study meets standards with distinction* is assigned when the average score is between 2.6 and 3.0. Chapter 5 provides the standards for cost analyses. Studies that include a cost-effectiveness and/or cost-benefit analysis will be reviewed and scored separately using the standards for each type of analysis, respectively, presented in Appendix E.

Appendix B: PRISMA-P and PRISMA-CI Elements

Exhibit B.1. PRISMA-P Elements

ID	Element	Explanation	Section addressing
1a	Title	The title of this report, *Protocol for the Pathways to Work Evidence Clearinghouse: Methods and Standards*, clearly identifies this as a review protocol.	Front matter
1b	Update	This review updates work done under the Employment Strategies for Low-Income Adults Evidence Review (ESER).	Appendix A
2	Registry	We were unable to register this review with the only logical registry (PROSPERO) because that review requires prospective registration. Pathways to Work partially relies on previously completed reviews conducted under ESER, so it was not possible to register Pathways to Work's standards before reviews began.	Not applicable
3a	Contact	Information on the authors and their institution appears on the title page.	Front matter
3b	Contributions	The ordering of the authors provides information on the relative contributions of each.	Front matter
4	Amendments	We will identify any future amendments by issuing an updated version of this document that clearly indicates the changes made.	Section 1.2, Appendix A
5a	Sources	This work was funded by the Office of Planning, Research, and Evaluation (OPRE), within the Administration for Children and Families (ACF), U.S. Department of Health and Human Services (HHS).	Front matter, Chapter 1
5b	Sponsor	This work was funded by OPRE, within ACF, HHS.	Front matter, Chapter 1
5c	Role of sponsor or funder	Staff from OPRE provided comments on and approved this protocol. They also helped shape the scope of the review.	Chapters 1 and 2
6	Rationale	The Pathways to Work Evidence Clearinghouse seeks to be a comprehensive resource that a range of audiences, including state and local TANF administrators, can use to identify the services that will best help people with low incomes succeed in the labor market.	Chapter 1
7	Objective	This review seeks to provide an overview of the broad field of the effectiveness of employment-related services and policies for people with low incomes. It aims to answer the following research questions: 1. What research exists on the effectiveness of programs that have the primary aim of improving the employment and earnings of people with low incomes? 2. Which programs and policies have evidence of improving employment, earnings, education, and training for people with low incomes and of reducing public benefit receipt?	Chapter 1
8	Eligibility criteria	Research must have met the following criteria: 1. Been published or prepared in 1990 or later 2. Conducted in the United States or Canada 3. Assessed effectiveness using quantitative methods	Chapter 2

ID	Element	Explanation	Section addressing
		4. Examined an intervention serving people with low incomes 5. Examined an intervention aiming to improve employment or earnings 6. Examined the impacts of an intervention on employment and/or earnings outcomes 7. Examined an intervention serving individual job-seekers in a specific context 8. Articulated details on the services provided	
9	Information sources	The review draws on a combination of database searches, literature reviews, a search of organizational websites, other federal review efforts, and a call for papers.	Section 2.2
10	Search strategy	The review used a modified version of the Peer Review of Electronic Search Strategies (PRESS) method (McGowen et al. 2016) to develop the database search terms in Exhibit 2.1 in addition to expert input on the search terms.	Section 2.2.3
11a	Data management	The project uses databases to catalog manuscripts and their corresponding studies as a management tool to track the literature search, screening, and review process.	Section 1.2
11b	Selection process	A two-stage screening process is used, and two reviewers examine each study.	Section 2.3; Section 3.3
11c	Data collection process	Data is recorded using a template based on that previously used by the ESER team. Pathways to Work Evidence Clearinghouse staff conduct author queries to gather information not reported in the study.	Section 3.3.1, Section 3.4
12	Data items	Team members collect data at the study, manuscript, finding, and program levels.	Section 3.4
13	Outcomes and prioritization	The Pathways to Work Evidence Clearinghouse team examines findings for outcomes in ten domains: short-term earnings, long-term earnings, very long-term earnings, short-term employment, long-term employment, very long-term employment, short-term public benefit receipt, long-term public benefit receipt, very long-term public benefit receipt, and education and training.	Section 3.2.2
14	Risk of bias in individual studies	Studies and findings are assigned a study quality rating based on several criteria. Findings are not reported for studies without sufficient causal validity.	Section 3.2
15	Synthesis	Studies are grouped into programs and findings are summarized by program. Future efforts will examine other groupings, potentially including meta-analysis.	Chapter 4
16	Meta-bias	This element will vary based on the syntheses conducted and will be elaborated upon in future synthesis briefs.	Not applicable
17	Confidence in cumulative evidence	The confidence in the evidence on each program is summarized by the intervention's rating.	Section 4.2

Note: This exhibit follows Moher et al. (2015)

Exhibit B.2. PRISMA-CI methods elements not discussed in PRISMA-P

ID	Element	Explanation	Section addressing
11a	Pathway complexity	This element will vary across programs. Pathways complexity will be elaborated in future synthesis briefs.	Not applicable
11b	Intervention complexity	This element will vary across programs and will be elaborated in an implementation brief for each program receiving a Well-Supported or Supported effectiveness rating. In these briefs, the Pathways to Work Evidence Clearinghouse team will detail available information on program components; the expected and actual frequency, duration, and intensity of service receipt; and the staff involved in service receipt.	Not applicable
11c	Population complexity	Studies examining people ages 16 and older with low incomes are eligible for review. Each study review further documents population characteristics.	Section 2.1; Section 3.4
11d	Implementation complexity	This element will vary across programs and will be elaborated in an implementation brief for each program receiving a Well-Supported or Supported effectiveness rating. In these briefs, the Pathways to Work Evidence Clearinghouse team will detail available information on key implementation drivers.	Not applicable
11e	Contextual complexity	This element will vary across programs and will be elaborated in an implementation brief for each program receiving a Well-Supported or Supported effectiveness rating. In these briefs, the Pathways to Work Evidence Clearinghouse team will detail available information on the location of service receipt and local context.	Not applicable
11f	Timing	Services and policies can occur for any length of time; however, the review restricts attention to analyses conducted in 1990 or later.	Chapter 2
13	Summary measures	We report effect sizes for each finding and average effect sizes by outcome domain and intervention.	Section 3.4; Section 4.2
14	Synthesis of results	Studies are grouped into programs and findings are summarized by program. Future efforts will examine other groupings, potentially including meta-analyses.	Chapter 4
16	Additional analyses	We will identify any additional analyses by issuing an updated version of this document, which clearly indicates the changes in a distinct section that summarizes updates.	Not applicable

Note: This exhibit follows Guise et al. (2017b)

Appendix C: Attrition Boundary

Exhibit C.1. Highest differential attrition rate for sample to maintain low attrition, by overall attrition rate

Overall	Differential	Overall	Differential	Overall	Differential
0	5.7	22	5.2	44	2.0
1	5.8	23	5.1	45	1.8
2	5.9	24	4.9	46	1.6
3	5.9	25	4.8	47	1.5
4	6.0	26	4.7	48	1.3
5	6.1	27	4.5	49	1.2
6	6.2	28	4.4	50	1.o
7	6.3	29	4.3	51	0.9
8	6.3	30	4.1	52	0.7
9	6.3	31	4.0	53	0.6
10	6.3	32	3.8	54	0.4
11	6.2	33	3.6	55	0.3
12	6.2	34	3.5	56	0.2
13	6.1	35	3.3	57	0
14	6.0	36	3.2	58	-
15	5.9	37	3.1	59	-
16	5.9	38	2.9	60	-
17	5.8	39	2.8	61	-
18	5.7	40	2.6	62	-
19	5.5	41	2.5	63	-
20	5.4	42	2.3	64	-
21	5.3	43	2.1	65	-

Source: What Works Clearinghouse Procedures and Standards Handbook, Version 5.0.

Appendix D: Details on Effect Size Calculation

The Pathways to Work Evidence Clearinghouse attempts to estimate effect sizes for each finding rated as providing high or moderate evidence. Specifically, Pathways to Work Evidence Clearinghouse team members calculate effect sizes as Hedges' *g*, defined as:

$$g = \frac{\omega(y_i - y_c)}{S},$$

where y_i and y_c are the means of the outcome for the intervention and comparison groups, ω is an adjustment for sample size, and *S* is the pooled standard deviation of the outcome. ω and *S* are further defined as

$$\omega = 1 - \frac{3}{4(n_i + n_c) - 9}$$

and

$$S = \sqrt{\frac{(n_i - 1)s_i^2 + (n_c - 1)s_c^2}{n_i + n_c - 2}},$$

where n_i and n_c are the number of people in the intervention and comparison groups, and s_i^2 and s_c^2 are the variances of the outcome for the intervention and comparison groups. When s_i^2 and s_c^2 are not both available, the Pathways to Work Evidence Clearinghouse team uses an alternative measure for *S* based on one of the group-specific measures (that is, $S = s_i$ or $S = s_c$) or a measure of the standard deviation of the outcome taken across the pooled intervention and comparison groups. These accommodations could cause small differences in effect sizes; however, they should not produce qualitatively different results.

The Pathways to Work Evidence Clearinghouse team uses methods for calculating Hedges' *g* that make the most use of information the study authors provide in the manuscripts under review (Exhibit D.1). This flexibility should minimize the burden on authors to provide supplemental information to the Clearinghouse.

As discussed in the body of the report, the Pathways to Work Evidence Clearinghouse team attempts to calculate two effect sizes for each finding receiving a high or moderate study quality rating (by finding): one that uses study-specific data to normalize the impact estimate, and one that uses national data to normalize the impact estimate. The team calculates an effect size using study-specific data if the Pathways to Work Evidence Clearinghouse can obtain from the study authors the information needed to apply one of the formulas in Exhibit D.1, either directly through the study, by calculating needed statistics based on the information reported in the study, or through an author query. The team calculates an effect size using national data when a nationally representative measure of the standard deviation of the outcome can be constructed.

Exhibit D.1. Alternative estimates of Hedges' *g*

Case	Method or formula Clearinghouse will use to calculate Hedges' *g*
The authors provide regression-adjusted means of the outcome (y_i' and y_c') and the information needed to estimate *S*.	$g = \frac{\omega(y_i' - y_c')}{S}$
The authors provide an impact estimate *(β)* from a regression.	$g = \frac{\omega\beta}{S}$
The authors provide a measure of effect size calculated using Hedges' *g (g')* that was not adjusted for sample size.	$g = \omega g'$
The authors provide a measure of effect size calculated using Cohen's *d* or Glass's delta.	These effect sizes use formulas similar to Hedges' *g*. The Pathways to Work Evidence Clearinghouse team will therefore use the measures provided, applying the sample size correction (ω) if needed.
The authors provide unadjusted means of the outcome (y_i and y_c) and the information needed to estimate *S*.	$g = \frac{\omega(y_i - y_c)}{S}$
The authors provide a *t*-statistic *(t)* from a simple test of differences in means or a regression without additional control variables.	$g = \omega t \sqrt{\frac{n_i + n_c}{n_i n_c}}$
The authors provide a *z*-statistic *(z)* from a simple test of differences in means or a regression without additional control variables.	$g = \omega z \sqrt{\frac{n_i + n_c}{n_i n_c}}$
The authors provide an *F*-statistic *(F)* from a simple test of differences in means or a regression without additional control variables.	$g = \omega \sqrt{\frac{F(n_i + n_c)}{n_i n_c}}$
The authors provide a *p*-value from a simple test of differences in means or a regression without additional control variables.	Use the t-distribution to determine the *t*-statistic associated with the provided *p*-value and calculate $g = \omega t \sqrt{\frac{n_i + n_c}{n_i n_c}}$
The authors provide the odds ratio *(OR)* for a binary outcome.	Hedges' *g* might not be estimated, but we can estimate a Cox index instead: $d_{Cox} = \omega \frac{ln(OR)}{1.65}$

Source: What Works Clearinghouse (WWC, 2022).
Note: Estimates of Glass' delta, Hedges' *g*, Cohen's *d*, and the Cox index can all be compared with one another. For examples see WWC (2022).

The Pathways to Work Evidence Clearinghouse team used data from the Integrated Public Use Microdata Series version of the Current Population Survey (CPS) to estimate nationally representative standard deviations of outcomes (Flood et al. 2018). The CPS is a nationally representative survey of U.S. households that has consistently collected information on income and employment since 1962. The team originally used the CPS to calculate one standard deviation for each outcome in each year from 1990 to 2018. The team will continue to update this information for subsequent years as data become available.

To estimate the appropriate standard deviations using national data, the team first identified the people in the CPS who could reasonably be considered to have "low income." Ideally, this would include people with low earnings potential and not those who have low earnings as the result of temporary investments in education or unemployment (for example, a graduate student pursuing an advanced degree, or a highly skilled individual who was recently laid off). To identify people with lower earnings potential, the team first ran a regression analysis using education, age, gender, and race and ethnicity to predict income within each CPS survey year from 1990 to 2018 (including only people ages 16 to 65). The team then defined people as having low income if their predicted income is in the bottom 20 percent of the distribution of predictions. The Pathways to Work Evidence Clearinghouse selected this threshold because about 20 percent of adults in the United States participate in government assistance programs in any given month (Irving and Loveless, 2015). Finally, the team used the actual outcome values for this population to estimate outcome standard deviations.

The Pathways to Work Evidence Clearinghouse used the CPS to calculate standard deviations for several key outcomes, listed below. The Pathways to Work Evidence Clearinghouse expects to select these outcomes most often for review:[33]

- Annual earnings (wage and salary income)
- Annual cash-based public assistance income
- Number of months received cash-based public assistance in past year
- Annual value of food stamps or Supplemental Nutrition Assistance Program (SNAP) benefits
- Monthly value of food stamps or SNAP benefits (only available from 1995 to 2014)
- Number of months received food stamps or SNAP benefits in past year
- Hourly wage rates
- Weekly earnings in current job

In addition, standard deviations for the following measures cannot be directly assessed using the CPS but can be estimated based on other CPS data using a few assumptions.

Monthly and quarterly earnings. The Pathways to Work Evidence Clearinghouse estimated the standard deviations of monthly and quarterly income using information from the CPS on the variance of annual income and the variance of weekly income, and one key assumption about how people's incomes vary over time. In particular, suppose that an individual's earnings in week i, x_i, follows the trajectory $x_i = \rho x_{i-1} + \varepsilon$, where ε is a random error and $0 < \rho < 1$, and that the variance of x_i is constant over the course of a year. Then the variance of income summed over multiple weeks can be written as

$$var(\sum_{i=1}^{N} x_i) = var(x_i)[N + 2\sum_{i=1}^{N}\sum_{j=i+1}^{N} corr(x_i, x_j)].$$

Because of the trajectory that income is assumed to follow, $corr(x_i, x_j) = \rho^{|i-j|}$. Therefore, this can be reduced to

[33] All listed outcomes are continuous. Standard deviations for binary outcomes can be calculated based on the means of these variables, making the use of nationally representative data unnecessary.

$$var(\sum_{i=1}^{N} x_i) = var(x_i)\left[N + 2\sum_{j=0}^{N-2}\sum_{i=1}^{N-j-1}\rho^i\right] = var(x_i)\left[N + 2\frac{\rho}{1-\rho}\left\{N - 1 - \left(\frac{\rho-\rho^N}{\rho^{-1}-1}\right)\right\}\right].$$

Then, if y_i is a person's annual income, it can be shown that

$$var(y_i) = var(x_i)\left[52 + 2\frac{\rho}{1-\rho}\left\{51 - \left(\frac{\rho-\rho^{52}}{\rho^{-1}-1}\right)\right\}\right].$$

This will enable the Pathways to Work Evidence Clearinghouse team to estimate ρ, which in turn will allow the team to estimate the variances of quarterly and monthly earnings, $var(\sum_{i=1}^{13} x_i)$ and $var(\sum_{i=1}^{4} x_i)$, respectively.

Monthly cash-based public assistance. The CPS data provide information on annual receipt of cash-based public assistance, as well as the number of months an individual received this assistance. The Pathways to Work Evidence Clearinghouse team estimated monthly income from public assistance by assuming that an individual received the same amount of cash assistance in each month that any such income was received over the course of the year. That is, if *n* is the number of months an individual received cash assistance, and *y* is the amount of assistance received in the past year, the team assumed monthly assistance received was *y/n* in months when any income was received and 0 otherwise. The team can then use the standard deviation of this measure to calculate effect sizes for outcomes measuring monthly cash-based public assistance income.

Monthly value of food stamps or SNAP benefits for 1990–1994 and 2015 onward. Although the CPS reports the monthly value of food stamp or SNAP benefits for 1995–2014, it does not contain this measure in other years. The Pathways to Work Evidence Clearinghouse team therefore assumes the ratio of the standard deviations of the annual and monthly values of food stamp or SNAP benefits remained constant from 1990–1995 and from 2014 onward.

Appendix E: Cost-Effectiveness and Cost-Benefit Analysis Standards

E.1 Cost-Effectiveness Analysis Standards

E.1.1 Cost-Effectiveness Analysis Threshold Standards

All studies will be reviewed as a cost analysis (See Chapter 5 of the main text). Studies that include a cost-effectiveness analysis will also be reviewed separately using standards specific to cost-effectiveness analyses. The threshold standards for a cost-effectiveness analysis are listed below.

For each threshold *standard*, reviewers choose "yes" or "no" based on whether a cost-effectiveness study meets the *criteria* defined for the standard. The standards are numbered, with each standard having a list of criteria (e.g., a, b, c) that must be satisfied for a standard to be met.

1. The cost effectiveness study describes the program of interest by satisfying the criteria listed below. The following three criteria must be satisfied:
    a. The study includes a description of the key program components as well as the context in which the program was implemented (e.g., specific geographic location of the program sites).
    b. The study includes a description of the program participants, including the number of participants and any available background characteristics (e.g., participant age, education level, socioeconomic status, income level, race/ethnicity, gender), preferably by site.
    c. Any study that includes a cost-effectiveness must include a description of the "business-as-usual condition" (BAU, i.e., what program participants would normally receive or experience in lieu of the intervention), including the program components.
2. The cost effectiveness study indicates that program costs are estimated using a **resource-based approach** that accounts for all of the **personnel resources** and **non-personnel resources** used for implementation. The following three criteria must be satisfied:
    a. The study demonstrates that a valid resource-based approach was used to estimate implementation costs (e.g., application of the **ingredients method**, a resource cost model).
    b. The study accounts for the full set of resources used to implement the program, including any resources that were purchased, redirected to the program from another purpose, or donated/volunteered.
    c. The study provides a description of the personnel and non-personnel resources accounted for in the cost estimate, including the quantities of the resources used and their associated prices.
3. The cost effectiveness study reports for which entities costs are being calculated (i.e., the analysis **perspective**).
    a. The study clearly states the perspective(s) from which costs are represented. This could be the societal perspective, the public perspective, the private perspective, or

the perspective of more granular groups, such as specific agencies, funders, or program participants.

4. The cost effectiveness study estimates and reports the per-participant cost of the program, including the site-specific per-participant costs and/or the average cost per participant across sites. The following two criteria must be satisfied:
    a. The study includes at least one of the following: (1) the site-specific per-participant cost for each site included in the cost study, or (2) the overall average per-participant cost of the program, including how the site-specific per-participant costs were used to estimate the average per-participant cost across sites (e.g., weighting by the number of participants served versus taking a straight average to estimate costs across sites).
    b. The study must report the difference between the per-participant costs of the program and the BAU condition, also known as the **incremental cost** of the program.
5. The outcome(s) used in a cost-effectiveness study are reported and align with at least one of the four Pathways to Work Evidence Clearinghouse outcome domains: earnings, employment, benefit receipt, and education and training. The following two criteria must be satisfied:
    a. The study states which of the four Pathways to Work Evidence Clearinghouse outcome domains are included in the study.
    b. The study reports the primary outcomes for which impacts were causally estimated and used to produce cost-effectiveness ratio(s).

E.1.2 Cost-Effectiveness Analysis Quality Rating Standards

Cost effectiveness studies that meet all threshold standards will then be evaluated by reviewers using the cost-effectiveness analysis quality rating standards, which provide additional information on the quality of the cost-effectiveness estimates. Each cost-effectiveness analysis quality rating standard is scored on a scale from 1 to 3, where 3 is the highest quality rating (some quality standards only have two possible rating scores of 1 or 3). At the end of the review, a summative cost-effectiveness analysis quality rating score is determined based on the average quality rating scores across all cost-effectiveness quality standards. The summative rating score is then rounded to the nearest tenth and converted into one of three cost study summative ratings:[34]

- **Cost Study Meets Standards With Low Quality**. Assigned for an average quality score between 1.0 and 1.5.
- **Cost Study Meets Standards With Moderate Quality**. Assigned for an average quality score between 1.6 and 2.5.
- **Cost Study Meets Standards With Distinction**. Assigned for an average quality score between 2.6 and 3.0.

Cost-Effectiveness Analysis Quality Rating Standards

[34] For example, summative rating scores of both 2.48 and 2.52 would both be rounded to 2.5 and receive a "Moderate Quality" rating; a score of 2.58 would be rounded to 2.6 and receive a "With Distinction" rating.

1. The cost-effectiveness study specifies whether the prices used to estimate costs reflect actual **local prices** or have been standardized to state or national average price levels. Scored as either 1 or 3, as follows:
    - (1) The study does not mention whether prices are local or represent standardized state or national prices.
    - (3) Prices are reported as being local or standardized to represent some other geographic area.
2. The cost-effectiveness study uses prices that represent the **opportunity cost** of the resources used, based on prices found in competitive markets or a justifiable **shadow price**. Scored as:
    - (1) The study provides no information on the prices used to calculate the costs of resources.
    - (2) The study includes the prices that were used to calculate resource costs.
    - (3) The study includes the prices that were used to calculate resource costs and lists the pricing sources.
3. The cost-effectiveness study uses prices that represent a common year for the resources included in the cost estimate. Scored as:
    - (1) The study provides no information on whether the resource prices represent a common year.
    - (2) The study includes information showing that prices used to calculate program costs represent a common year.
    - (3) The study includes information showing that prices used to calculate program costs represent a common year and describes the inflation adjustments used to bring all prices to a common year.
4. The cost-effectiveness study reports the present value of program costs across the analysis period and expresses these for a common reference year. Scored as either 1 or 3, as follows:
    - (1) The study does not report at least one of the following components needed to calculate a present value: (a) the specific year for which costs are represented, (b) the period over which program costs were incurred, and (c) the discount rate used in the calculation of present values.
    - (3) The study reports all of the following components needed to calculate a present value: (a) the specific year for which costs are represented, (b) the period over which program costs were incurred, and (c) the discount rate used in the calculation of present values.
5. The sites included in the cost-effectiveness study are representative of those in which the program of interest is being implemented and the sites used to estimate impacts supporting the analyses. Scored as:
    - (1) The study does not include information about site selection for the cost-effectiveness study.

- (2) The study includes information about the sites included in the cost-effectiveness study but does not explain whether that selection is representative of the sites in which the program was implemented and/or representative of the sites used to estimate impacts.
- (3) The study provides evidence that the sites included in the cost-effectiveness study are representative of the sites in which the program was studied and the sample of sites used to estimate impacts, or that all sites were included in the cost study. Evidence of representativeness is demonstrated by showing similarities between the analysis program sites and population of sites with respect to any of the following characteristics: background characteristics of participants (e.g., age, education level, socioeconomic status, income level, race/ethnicity, gender), as well as the sites' geographic region and locale (e.g., urban versus rural).

6. The cost-effectiveness study specifies when cost data collection occurred relative to the period when program costs were incurred (i.e., **concurrent**, **retrospective**, or **prospective data collection**). Scored as:
   - (1) The study does not include details about when information to estimate costs was collected.
   - (2) The study indicates that the information used to estimate costs was collected prospectively or retrospectively to when program costs were incurred.
   - (3) The study indicates that information used to estimate costs was collected concurrently with when the program costs were incurred.
7. The cost-effectiveness study reports how program resources are financed or otherwise provided by different entities and how costs are distributed across entities. Examples of entities include program participants, program staff, program funders, or resource providers. In addition to being directly financed, new **resources** may be provided through volunteer time and/or donated non-personnel items, whereas existing resources may be redirected from other programs or purposes. Scored as:
   - (1) The study has no description of how resources were financed or otherwise provided by different entities nor how costs were distributed across different entities.
   - (2) The study either describes how program resources were financed or otherwise provided by different entities, or describes how costs were distributed across entities, but not both.
   - (3) The study describes both how program resources were financed or otherwise provided by different entities, and how costs were distributed across entities.
8. The cost-effectiveness study reports the findings of a **sensitivity analysis** that evaluates how results change in response to alternative assumptions used to estimate costs and effectiveness. Assumptions may relate to types or quantities of resources used and their prices, magnitude of outcome impact, characteristics of the participants served, sites included in the analysis, discount rates used to calculate present values, the time horizons over which costs are calculated, or other relevant assumptions. Scored as:
   - (1) The study has no mention of a cost study sensitivity analysis being conducted.

- (2) The study mentions that a cost study sensitivity analysis was conducted but does not provide a detailed description of the findings of the analyses.
- (3) The study provides an in-depth description of least one cost study sensitivity analysis that was conducted, including what assumptions were explored and the degree to which the study findings were sensitive to these assumptions.

9. The cost-effectiveness study describes whether key program resources correspond to **fixed costs**, **variable costs**, or **lumpy costs**, for the purpose of describing the cost implications of implementation at different scales. Scored as:
    - (1) The report does not include information about whether the key resources used to implement the program are associated with fixed, variable, or lumpy costs.
    - (2) The study includes information about whether the key resources used are associated with fixed, variable, or lumpy costs but does not use this information to discuss the cost implications of implementation at different scales.
    - (3) The study includes information about key resources used being related to fixed, variable, or lumpy costs and uses this information to discuss the cost implications of implementation at different scales.
10. A cost-effectiveness analysis either qualitatively or quantitatively reports the costs associated with services that are induced by participation in the program of interest but are delivered outside of the program itself (treated as **induced costs** in a cost-effectiveness analysis). Scored as:
    - (1) The study includes no mention of induced costs or negative benefits.
    - (2) The study mentions induced costs but does not describe (neither qualitatively nor quantitatively) what those were.
    - (3) The study describes, either qualitatively or quantitatively, the induced costs associated with program participation.

E.2 Cost-Benefit Analysis Standards

E.2.1 Cost-Benefit Analysis Threshold Standards

All studies will be reviewed as a cost analysis (See Chapter 5 of the main text). Studies that include a cost-benefit analysis will also be reviewed separately using standards specific to cost-benefit analyses. The threshold standards for a cost-benefit analysis are listed below.

For each cost-benefit analysis threshold *standard*, reviewers choose "yes" or "no" based on whether a cost-benefit study meets the *criteria* defined for the standard. The standards are numbered, with each standard having a list of criteria (e.g., a, b, c) that must be satisfied for a standard to be met.

1. The cost-benefit study describes the program of interest by satisfying the criteria listed below. The following three criteria must be satisfied:
    a. The study includes a description of the key program components as well as the context in which the program was implemented (e.g., specific geographic location of the program sites).

b. The study includes a description of the program participants, including the number of participants and any available background characteristics (e.g., participant age, education level, socioeconomic status, income level, race/ethnicity, gender), preferably by site.

c. A cost-benefit analysis must include a description of the "business-as-usual condition" (BAU, i.e., what program participants would normally receive or experience in lieu of the intervention), including the program components.

2. The cost-benefit study indicates that program costs are estimated using a **resource-based approach** that accounts for all of the **personnel resources** and **non-personnel resources** used for implementation. The following three criteria must be satisfied:

    a. The study demonstrates that a valid resource-based approach was used to estimate implementation costs (e.g., application of the **ingredients method**, a resource cost model).

    b. The study accounts for the full set of resources used to implement the program, including any resources that were purchased, redirected to the program from another purpose, or donated/volunteered.

    c. The study provides a description of the personnel and non-personnel resources accounted for in the cost estimate, including the quantities of the resources used and their associated prices.

3. The cost-benefit study reports for which entities costs and benefits are being calculated (i.e., the analysis **perspective**).

    a. The study clearly states the perspective(s) from which costs and benefits are represented. This could be the societal perspective, the public perspective, the private perspective, or the perspective of more granular groups, such as specific agencies, funders, or program participants.

4. The cost-benefit study estimates and reports the per-participant cost of the program, including the site-specific per-participant costs and/or the average cost per participant across sites. The following two criteria must be satisfied:

    a. The study includes at least one of the following: (1) the site-specific per-participant cost for each site included in the cost study, or (2) the overall average per-participant cost of the program, including how the site-specific per-participant costs were used to estimate the average per-participant cost across sites (e.g., weighting by the number of participants served versus taking a straight average to estimate costs across sites).

    b. A cost-benefit analysis must report the difference between the per-participant costs of the program and the BAU condition, also known as the **incremental cost** of the program.

5. The outcome(s) used in a cost-benefit analysis are reported and align with at least one of the four Pathways to Work Evidence Clearinghouse outcome domains: earnings, employment, benefit receipt, and education and training. The following criteria must be satisfied:

    a. The study states which of the four Pathways to Work Evidence Clearinghouse outcome domains are included in the study.

b. A cost-benefit analysis reports the primary outcomes for which impacts were causally estimated, along with any related outcomes that were used to calculate benefits.

E.2.2 Cost-Benefit Analysis Quality Rating Standards

Cost-benefit studies that meet all threshold standards will then be evaluated by reviewers using the cost-benefit analysis quality rating standards, which provide additional information on the quality of the cost-benefit estimates. Each applicable quality rating standard is scored on a scale from 1 to 3, where 3 is the highest quality rating (some quality standards only having two ratings scores of 1 or 3). At the end of the review, a summative cost-benefit analysis quality rating score is determined based on the average quality rating scores across all cost-benefit quality standards. The summative rating score is then rounded to the nearest tenth and converted into one of three cost study summative ratings:[35]

- **Cost Study Meets Standards With Low Quality**. Assigned for an average quality score between 1.0 and 1.5.
- **Cost Study Meets Standards With Moderate Quality**. Assigned for an average quality score between 1.6 and 2.5.
- **Cost Study Meets Standards With Distinction**. Assigned for an average quality score between 2.6 and 3.0.

Cost-Benefit Analysis Quality Rating Standards

1. The cost-benefit study specifies whether the prices used to estimate costs reflect actual **local prices** or have been standardized to state or national average price levels. Scored as either 1 or 3, as follows:
    - (1) The study does not mention whether prices are local or represent standardized state or national prices.
    - (3) Prices are reported as being local or standardized to represent some other geographic area.
2. The cost-benefit study uses prices that represent the **opportunity cost** of the resources used, based on prices found in competitive markets or a justifiable **shadow price**. Scored as:
    - (1) The study provides no information on the prices used to calculate the costs of resources.
    - (2) The study includes the prices that were used to calculate resource costs.
    - (3) The study includes the prices that were used to calculate resource costs and lists the pricing sources.
3. The cost-benefit study uses prices that represent a common year for the resources included in the cost estimate. Scored as:

[35] For example, summative rating scores of both 2.48 and 2.52 would both be rounded to 2.5 and receive a "Moderate Quality" rating; a score of 2.58 would be rounded to 2.6 and receive a "With Distinction" rating.

- (1) The study provides no information on whether the resource prices represent a common year.
- (2) The study includes information showing that prices used to calculate program costs represent a common year.
- (3) The study includes information showing that prices used to calculate program costs represent a common year and describes the inflation adjustments used to bring all prices to a common year.

4. The cost-benefit study reports the present value of program costs and benefits across the analysis period and expresses these for a common reference year. Scored as either 1 or 3, as follows:
    - (1) The study does not report at least one of the following components needed to calculate a present value: (a) the specific year for which costs and benefits are represented, (b) the period over which program costs were incurred and benefits accrued, and (c) the discount rate used in the calculation of present values.
    - (3) The study reports all of the following components needed to calculate a present value: (a) the specific year for which costs and benefits are represented, (b) the period over which program costs were incurred and benefits accrued, and (c) the discount rate used in the calculation of present values.
5. The sites included in the cost-benefit study are representative of those in which the program of interest is being implemented and the sites used to estimate impacts supporting the analyses. Scored as:
    - (1) The study does not include information about site selection for the cost-benefit study.
    - (2) The study includes information about the sites included in the cost-benefit study but does not explain whether that selection is representative of the sites in which the program was implemented and is representative of the sites used to estimate impacts.
    - (3) The study provides evidence that the sites included in the cost-benefit study are representative of the sites in which the program was studied and the sample of sites used to estimate impacts, or that all sites were included in the cost study. Evidence of representativeness is demonstrated by showing similarities between the analysis program sites and population of sites with respect to any of the following characteristics: background characteristics of participants (e.g., age, education level, socioeconomic status, income level, race/ethnicity, gender), as well as the sites' geographic region and locale (e.g., urban versus rural).
6. The cost-benefit study specifies when cost data collection occurred relative to the period when program costs were incurred (i.e., **concurrent**, **retrospective**, or **prospective data collection**). Scored as:
    - (1) The study does not include details about when information to estimate costs was collected.
    - (2) The study indicates that the information used to estimate costs was collected prospectively or retrospectively to when program costs were incurred.

- (3) The study indicates that information used to estimate costs was collected concurrently with when the program costs were incurred.

7. The cost-benefit study reports how program resources are financed or otherwise provided by different entities and how costs and benefits are distributed across entities. Examples of entities include program participants, program staff, program funders, or resource providers. In addition to being directly financed, new **resources** may be provided through volunteer time and/or donated non-personnel items, whereas existing resources may be redirected from other programs or purposes. Scored as:
    - (1) The study has no description of how resources were financed or otherwise provided by different entities nor how costs and benefits were distributed across different entities.
    - (2) The study either describes how program resources were financed or otherwise provided by different entities, or describes how costs and benefits were distributed across entities, but not both.
    - (3) The study describes both how program resources were financed or otherwise provided by different entities, and how costs and benefits were distributed across entities.
8. The cost-benefit study reports the findings of a **sensitivity analysis** that evaluates how results change in response to alternative assumptions used to estimate costs, effectiveness, or benefits. Assumptions may relate to types or quantities of resources used and their prices, magnitude of outcome impact, characteristics of the participants served, sites included in the analysis, discount rates used to calculate present values, the time horizons over which costs (or benefits) are calculated, or other relevant assumptions. Scored as:
    - (1) The study has no mention of a cost study sensitivity analysis being conducted.
    - (2) The study mentions that a cost study sensitivity analysis was conducted but does not provide a detailed description of the findings of the analyses.
    - (3) The study provides an in-depth description of least one cost study sensitivity analysis that was conducted, including what assumptions were explored and the degree to which the study findings were sensitive to these assumptions.
9. The cost-benefit study describes whether key program resources correspond to **fixed costs**, **variable costs**, or **lumpy costs**, for the purpose of describing the cost implications of implementation at different scales. Scored as:
    - (1) The report does not include information about whether the key resources used to implement the program are associated with fixed, variable, or lumpy costs.
    - (2) The study includes information about whether the key resources used are associated with fixed, variable, or lumpy costs but does not use this information to discuss the cost implications of implementation at different scales.
    - (3) The study includes information about key resources used being related to fixed, variable, or lumpy costs and uses this information to discuss the cost implications of implementation at different scales.

10. A cost-benefit analysis either qualitatively or quantitatively reports the costs associated with services that are induced by participation in the program of interest but are delivered outside of the program itself (treated as **negative benefits** in a cost-benefit analysis). Scored as:
    - (1) The study includes no mention of induced costs or negative benefits.
    - (2) The study mentions negative benefits but does not describe (neither qualitatively nor quantitatively) what those were.
    - (3) The study describes, either qualitatively or quantitatively, the negative benefits associated with program participation.
11. A cost-benefit analysis reports the evaluation **perspective** being used, which should be the same for the estimated costs and benefits. Examples of evaluation perspectives include societal, public, and participant. Scored as either 1 or 3, as follows:
    - (1) The study indicates that different perspectives were used to calculate costs and benefits.
    - (3) The study indicates that the same perspective was used to calculate both costs and benefits.

Appendix F: Glossary of Cost Study Terms

Concurrent Data Collection: Occurs when cost analysis data collection happens concurrently with, or at the same time as, program implementation. This facilitates the capture of program costs as they are incurred and is likely to produce the most accurate estimate of costs.

Cost Analysis (CA): A systematic method for identifying and documenting the quantity, quality, and economic value, as represented by opportunity costs, of all resources such as personnel, materials, equipment, and facilities used to implement and operate a program. The analysis often categorizes costs according to resource type, program component, implementation phase and investigates the sensitivity of cost estimates by varying key assumptions.

Cost-Benefit Analysis (CBA): A type of cost study that compares the present discounted monetary values of intervention costs and outcomes (also referred to as benefit cost analysis).

Cost-Effectiveness Analysis (CEA): A comparative analysis in which two or more alternative means of obtaining the same objective are assessed based on their costs per unit of effectiveness. Because CEA is comparative, a program can be considered cost-effective only relative to something else.

Fixed Costs: Costs that do not depend on the number of participants (e.g., a program manager may spend 5% of their time supervising a program regardless of the number of participants).

Incremental (or Differential) Costs: Costs of program implementation and operation above or below the costs of an alternative program or condition. Because incremental costs are relative, they can be positive or negative (decremental).

Induced Costs (aka External or Mediated Costs): Costs arising from behavioral change after an intervention has been implemented that result in additional resources to be expended (Levin et al., 2018, pp. 52, 201 & 225). These include spillovers and additional costs incurred as a result of an intervention. For example, an intervention may have the positive effect of more students going to college, but this incurs costs to the students for tuition and foregone earnings and to the government in terms of financial aid and subsidies. In CA or CEA, these must be added to program costs. In CBA, they are counted as *negative benefits*.

Ingredients Method (aka Resource Cost Model): A method of accurately and consistently estimating the costs of a program through identification, quantification, and pricing of the comprehensive collection of personnel and non-personnel resources used for its implementation.

Local Prices: Prices of goods in the context of the local market that may differ from national prices. For example, program staff wages in one local context could vary significantly from a national average wage.

Lumpy Costs: Costs that increase in steps at predictable thresholds (e.g., an additional case manager may be needed for every 73rd participant).

Negative Benefits: Costs arising from behavioral change after an intervention has been implemented that result in additional resources to be expended (Levin et al., 2018, pp. 52, 201 & 225). These include spillovers and additional costs incurred as a result of an intervention. For example, an intervention may have the positive effect of more students going to college, but this incurs costs to the students for tuition and foregone earnings and to the government in terms of

financial aid and subsidies. In CA and CEA, they are considered to be *induced costs* and added to program costs. In CBA, these indirect costs are subtracted from the calculated benefits.

Non-personnel Resources: *Non-staff resources* that are used to implement a program including those that are: 1) purchased explicitly for the program, 2) existing resources that are redirected to the program from another purpose, or 3) donated to the program. Office space, utilities, supplies, and materials are examples of non-personnel resources.

Opportunity Cost: "The value of what is sacrificed by using a specific resource in one way rather than in its best alternative use" (Levin and Belfield, 2015, p. 403). In practice, this often means assigning a market price to a resource (e.g., salary and fringe benefits for personnel).

Outcome: Consequence of program implementation as measured by changes in behavior, performance, or other measure of interest.

Personnel Resources: Staff or other human resources used to implement an intervention. This includes all paid full-time and part-time staff, as well unpaid volunteers and participants (paid or unpaid) involved in implementing a program.

Perspective: The point of view taken when defining costs and benefits included in a study. Common perspectives taken include:

- **Participants' Perspective**: Includes only the costs and benefits borne by program participants and sometimes their families (e.g., the opportunity costs of time for students attending college, who could otherwise be in the workforce).
- **Program Provider Perspective**: The subset of costs that accrue to the organization implementing a program.
- **Societal Perspective**: Takes into account the opportunity costs of all resources required to implement and operate a program, regardless of who pays for or contributes the resources. This perspective also accounts for program benefits to all stakeholders.

Prospective Data Collection: Occurs when program costs are estimated in advance of program implementation. This can inform strategic planning or allow time for financing to be raised before implementation.

Retrospective Data Collection: Occurs when data is collected after a program has already been implemented and is no longer in operation. This can inform decisions about restarting or replicating a previous program, and can provide cost data that line up with past performance data.

Resource-based Approach to Estimate Costs: An approach that accounts for the comprehensive collection of the full set of resources used to implement the program, including all personnel (staff and personnel resources, including the quantities of the resources used, their associated prices, and sources of this information) and non-personnel resources (including any that were purchased, redirected to the program from another purpose, or donated/volunteered).

See also, *Ingredients Method (aka Resource Cost Model).*

Resources: Tangible inputs, such as personnel, materials, equipment, and facilities (physical spaces).

Sensitivity Analysis: An analysis varying the assumptions or parameter values used in the primary analysis to determine sensitivity or robustness of results to those assumptions.

Shadow Price: An estimate of an economic value when market-based values are unavailable (Crowley et al., 2018, p. 378).

Site: The location where a program is implemented and most information on resource use is gathered. For example, the site can be a classroom, point of service, county, or state.

State or National Average Prices: Prices of goods that are a state or national average (in contrast to local prices). Used when analysis combines data from multiple geographic locations, to value resources from all localities. National prices are often used as the reference case.

Time Horizon: The total period during which benefit and cost streams are either observed or predicted.

Variable Costs: Costs that vary proportionally with the number of program participants (e.g., the costs of individual participant workbooks) or the number of sites implementing the program.