



EMPLOYMENT STRATEGIES
for Low-Income Adults *Evidence Review*

OPRE REPORT 2015-27

Employment Strategies for Low-
Income Adults Evidence Review:
Standards and Methods

May 2015

Employment Strategies for Low-Income Adults Evidence Review: Standards and Methods

May 2015

Annalisa Mastri
Emily Sama-Miller
Andrew Clarkwest

Submitted to:

Office of Planning, Research, and Evaluation
Administration for Children and Families
U.S. Department of Health and Human Services
Project Officers: Emily Schmitt and Kimberly Clum
Contract Number: HHSP23320095642WC/HHSP2337044T

Submitted by:

Mathematica Policy Research
P.O. Box 2393
Princeton, NJ 08543-3535
Telephone: (609) 799-3535
Facsimile: (609) 799-0005
Project Director: Annalisa Mastri
Reference Number: 40307

This report is in the public domain. Permission to reproduce is not necessary. Suggested citation: Mastri, A., Sama-Miller, E., and Clarkwest, A. (2015). *Employment Strategies for Low-Income Adults Evidence Review: Standards and Methods*. Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. Washington, DC.

Disclaimer: The views expressed in this publication do not necessarily reflect the views or policies of the Office of Planning, Research and Evaluation, the Administration for Children and Families, or the U.S. Department of Health and Human Services.

This report and other reports sponsored by the Office of Planning, Research and Evaluation are available at <http://www.acf.hhs.gov/programs/opre/index.html>.

 CONTENTS

OVERVIEW	iv
INTRODUCTION.....	1
I. CONDUCTING A SYSTEMATIC, BROAD LITERATURE SEARCH.....	1
A. Examining existing literature reviews	2
B. Conducting database searches	2
C. Call for papers	3
D. Selected additions to the literature pool	3
II. SCREENING STUDIES AGAINST ELIGIBILITY CRITERIA	4
III. ASSESSING THE STRENGTH OF EVIDENCE	5
A. Standards	5
1. Eligible study designs.....	6
2. Overview of ratings, their interpretation, and relevant criteria	6
3. Assessing randomization procedures	9
4. Examining attrition in RCTs	9
5. Assessing equivalence on baseline characteristics for RCTs with high attrition and comparison group designs	11
6. Validity of outcome measures	12
B. Review process	13
C. Conflicts of interest	13
IV. EXTRACTING AND PRESENTING DATA	13
V. CONCLUSION	15
APPENDIX A: LITERATURE REVIEWS	A.1
APPENDIX B: CALL FOR PAPERS	B.1

TABLES

I.1	Keywords used in database search	2
I.2	Organizational websites included in custom Google search	3
III.1	Ratings and their interpretation	6
III.2	Highest differential attrition rate for sample to maintain low attrition, by overall attrition rate	10
IV.1	Information captured at the study and outcome level during study reviews	14
IV.2	Decision rules for selecting outcomes to report within each domain for each data source	16

OVERVIEW

The Employment Strategies for Low-Income Adults Evidence Review systematically reviewed the literature on the effectiveness of employment and training programs for low-income adults. Mathematica Policy Research conducted the review, under contract to the Office of Planning, Research, and Evaluation (OPRE) in the Administration for Children and Families (ACF). The purpose of the review was:

To provide practitioners, policymakers, researchers, and the general public with a transparent and systematic assessment of the research evidence for effectiveness of programs designed to improve the employment-related outcomes of low-income adults.

This report outlines the standards and methods that governed the review, under four main activities: searching the literature, screening studies for eligibility, assessing each study's strength of evidence, and extracting information reported in the studies.

We used a four-pronged approach to review the literature:

- Conducting a systematic, broad literature search to identify relevant studies
- Screening studies to determine whether they met the criteria for being reviewed
- Assessing the strength of evidence in the studies
- Extracting data and presenting studies' key findings to the review's audiences

To be eligible for review, a study must have:

- Quantitatively measured the effectiveness of a program using a design that compared the outcomes in a treatment group (that could receive the program) to a similar comparison group (that could not receive the program)
- Been published since 1990
- Aimed to improve employment-related outcomes
- Examined a program that served low-income adults
- Examined a program implemented in the United States, Canada, or the United Kingdom

The project reviewed studies to assess the strength of the evidence they presented for the effectiveness of a program (which could consist of an intervention, strategy, approach, or combination thereof).

- “Strength of evidence” refers to how likely it is that a study's estimated program impacts were *caused by* that program, not something else.
- The review assigned a high, moderate, or low rating to each study reviewed. To determine the rating, reviewers used a comprehensive set of review standards.
- The standards focused on whether the treatment group and the comparison group were similar before the program began. The two groups must have been either formed by a random process or shown to be similar on several pre-program characteristics in order for a study to rate high or moderate.

The review presents study findings on a project website that reports selected results for all eligible studies. The website also allows users to search for results by program studied, by target population, and in other ways.

INTRODUCTION

The Employment Strategies for Low-Income Adults Evidence Review systematically reviewed the literature on the effectiveness of employment and training programs for low-income adults. An employment program could consist of an intervention, strategy, approach, or combination thereof. Mathematica Policy Research conducted the review, under contract to the Office of Planning, Research, and Evaluation (OPRE) in the Administration for Children and Families (ACF). The purpose of the review was:

To provide practitioners, policymakers, researchers, and the general public with a transparent and systematic assessment of the research evidence for effectiveness of programs designed to improve the employment-related outcomes of low-income adults.

To be eligible for the review, studies had to meet three criteria that followed from the above statement of purpose:

1. Measured effectiveness using a quantitative approach
2. Examined a program intended to affect employment-related outcomes
3. Examined a program that served low-income adults

To further ensure the review's relevance to its intended audiences, the studies had to be published in 1990 or later and, with certain exceptions for the United Kingdom and Canada discussed in section II of this report, they must have focused on a program tested in the United States.

The review process included four major activities:

1. Conducting a systematic, broad literature search to identify relevant studies
2. Screening studies to determine whether they met the criteria for being included in the review
3. Assessing the strength of evidence in the studies
4. Extracting data and presenting studies' key findings to the review's audiences

The review's standards and procedures, described in this report, were developed collaboratively by OPRE and Mathematica, with OPRE giving final approval. A panel of expert consultants, including researchers and practitioners recognized for their experience with and knowledge of programs serving low-income individuals, as well as methodology experts, also contributed.

I. CONDUCTING A SYSTEMATIC, BROAD LITERATURE SEARCH

The review team identified potentially eligible studies in four steps. First, it examined reference lists in existing literature reviews, which ensured that the review would capture well-known studies. Second, to identify other relevant published literature, the team searched electronic citation databases by systematically using search terms relevant to the review's eligibility criteria. Third, the team circulated a call for papers which (1) helped to ensure that unpublished relevant studies could be included in the review and (2) reduced the risk of

publication bias, which occurs if studies that do not show impacts, however valuable otherwise, are less likely to be published. Finally, when examining the literature resulting from the first three steps, the team manually searched for and added related publications.

A. Examining existing literature reviews

Mathematica identified reviews of the literature on the effectiveness of employment programs (Appendix A) and submitted the list to OPRE. After OPRE approved the list, Mathematica examined the lists of references in the literature reviews and identified studies that were likely to meet the eligibility criteria.

B. Conducting database searches

The review team systematically searched electronic citation databases using keywords that reflected the scope of the review (Table I.1). Wherever possible, the keyword search examined studies' titles and abstracts, identified all studies that had at least one search term aligned with each of the eligibility criteria, and filtered the results to show only publications from 1990 through April 2014. Specifically, the review team searched using OR to connect keywords within each row of Table I.1, and AND to connect the eligibility criteria across rows. The review team included the following electronic citation databases in the keyword search: Academic Search Premier, EconLit, Education Research Complete, E-Journals, ERIC, PsycINFO, ProQuest Dissertations and Theses, Scopus, and SocINDEX with full text.

Table I.1. Keywords used in database search

Eligibility criterion	Keywords
Design terms	Effect*, causal, impact
Outcome terms	Employment, job, occupation, earnings, wage*
Sample terms	Low-income, disadvantaged, TANF, homeless, ex-offender, disabled
Geography terms	United States, U.S., US, state names, District of Columbia

Note: An asterisk indicates a truncation. When used in a search term, all words with the root will be returned. 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."

The team also performed keyword searches of research clearinghouses and working paper databases housed on publicly available websites: The Campbell Collaboration, NBER Working Papers, RePEc, Self-Sufficiency Research Clearinghouse, and Social Science Research Network. Finally, the team performed a custom Google search of the organizational websites listed in Table I.2 to identify additional studies. These sources of research were relevant to the review, but had restrictions such as not allowing search limitations by date range or restrictions to certain fields, so the team searched in a way that matched, as closely as possible, the process described above.

Table I.2. Organizational websites included in custom Google search

Organization	URL
Abdul Latif Jameel Poverty Action Lab	www.povertyactionlab.org/
Abt Associates	http://www.abtassociates.com/
Administration for Children and Families	https://www.acf.hhs.gov/
Center for Law and Social Policy	http://www.clasp.org/
Center on Poverty, Work and Opportunity	http://www.law.unc.edu/centers/poverty/
Center for Study of Urban Poverty	http://www.csup.ucla.edu/
Employment & Training Administration Research Database	http://wdr.doleta.gov/research/eta_default.cfm
Multidisciplinary Program in Inequality and Social Policy	http://inequality.hks.harvard.edu/
Impaq Associates	http://www.impaqint.com/
IZA	http://www.iza.org/en/webcontent/index_html
Institute for Research on Poverty	www.irp.wisc.edu/
Joblessness and Urban Poverty Research Program	http://www.hks.harvard.edu/urbanpoverty/
Joint Center for Poverty Research	www.jcpr.org/
Mathematica Policy Research	http://www.mathematica-mpr.com/
MDRC	www.mdrc.org/
National Center for Children in Poverty	www.nccp.org/
National Center for Policy Analysis	www.ncpa.org/
National Poverty Center	www.npc.umich.edu/
NORC	www.norc.org/
Ray Marshall Center	http://raymarshallcenter.org/
National Research Center for Women and Families (now known as National Center for Health Research)	http://center4research.org/
RAND	http://www.rand.org/
RTI International	http://www.rti.org/
Social Policy Research Associates	http://www.spra.com/
The Stanford Center on Poverty and Inequality	http://web.stanford.edu/group/scspi/
University of Kentucky Center for Poverty Research	http://www.ukcpr.org/
Upjohn Institute	http://www.upjohn.org/
Urban Institute	http://www.urban.org/

C. Call for papers

To identify studies that were not publicly available, the review team emailed a call for papers to research organizations, professional associations, and individual researchers (Appendix B). Mathematica and OPRE also publicized the call for papers through their respective social media accounts and newsletters. The team circulated the first call for papers on May 5, 2014, and the deadline to respond was June 17, 2014. To further ensure that all authors who wanted to respond could do so, the team recirculated the call for papers on June 25, 2014, and the deadline to respond was July 17, 2014.

D. Selected additions to the literature pool

After identifying a pool of literature through the first three steps, the team began screening the pool for review. During that process, the review team occasionally recognized gaps in the identified literature. For instance, the searches sometimes located an implementation report but not an impact report about the same evaluation. This was most common in grey literature, which is produced by government agencies and research organizations and is a common information source for the review. When such omissions were evident, the team used the studies the review team identified to locate and add related studies to the pool for screening and review (a method sometimes called snowballing).

II. SCREENING STUDIES AGAINST ELIGIBILITY CRITERIA

In reviewing the literature, the review team considered evidence of effectiveness for a range of programs to improve employment-related outcomes, including direct employment services such as job search assistance or training support. The team applied seven eligibility criteria in screening studies for the review. The first three followed directly from the review’s purpose. The next four, established through discussions with OPRE and expert consultants, were intended to ensure that the review focused on research that is most applicable in making policy decisions.

1. **Assessed effectiveness using quantitative methods.** This criterion screened out studies that used purely descriptive methods (for example, those that described employment trends among low-income adults) and studies that focused only on how a program was implemented. Effectiveness studies also had to be based on a design for which OPRE and the project team had identified review standards: randomized controlled trials, comparison group designs, or regression discontinuity designs.¹
2. **Examined a program with a primary aim of improving employment-related outcomes.**² An example of such a program is an occupational training program designed to improve participants’ employment-related outcomes. However, general education programs did not satisfy the “primary aim” criterion; although most such programs *could* affect employment-related outcomes, their educational content is neither tied to an occupation nor part of a broader program with other explicit employment supports.

The employment-related outcomes that were the focus of the review fall into seven domains: short- and long-term employment, short- and long-term earnings, short- and long-term public benefit receipt, and acquisition of education and/or training. Short-term was defined as 18 months or less.³ Studies had to examine program impacts in the employment or earnings domains to be included, though outcomes about public benefit receipt and education were also assessed and summarized by the review.

3. **Served low-income adults.** The review considered certain specific groups (homeless people, formerly incarcerated people, and public benefit recipients) to be generally low income. If the study did not focus on these groups, the author must have declared or shown the study sample to be low income for it to be eligible for review. The review focused on programs intended to serve individuals age 18 or older at the time of enrollment.
4. **Served individual job-seekers.** Studies that examined policies or actions that affected communities, such as enterprise zones, or employers, such as tax credits for hiring disadvantaged workers, were not eligible for the review.

¹ Standards for regression discontinuity designs are not discussed in this document because the literature search did not identify eligible studies that used such a design. Instrumental variables designs were not eligible for review because they were uncommon in this literature, and no comprehensive standards have been established.

² Rarely, approaches intended to improve employment outcomes by addressing other barriers to employment, such as health or housing problems, were also reviewed. This occurred only if these approaches are examined within the context of a larger evaluation that met the screening criteria for the review.

³ The reference point for defining 18 months was set by study authors. Some authors reported follow-ups relative to random assignment and others reported follow-ups relative to the beginning or end of program services.

5. **Was a defined, replicable employment program.** Eligible studies articulated the services job seekers received so that others could implement them. Research on a state’s welfare-to-work program met this criterion because such programs typically consist of specific combinations of strategies such as job search assistance and financial incentives that would be replicable. But national studies of “welfare reform” were not eligible because states had substantial flexibility in implementing welfare reform, so these studies did not examine a defined and replicable program.
6. **Published or prepared in 1990 or later.** This refers to the date of first publication (or, if submitted through the call for studies, unpublished manuscript) of results from an evaluation. A single evaluation may have had multiple publications presenting its results. Publications in or after 1990 that repackaged study conclusions disseminated before 1990 (such as a journal article summarizing an earlier report) were also ineligible.
7. **Conducted in the United States, or conducted in the United Kingdom or Canada and in a consulted literature review.** The review was designed to serve practitioners, policymakers, researchers, and the general public in the United States, so it focused on studies of programs tested in the United States. But the United States’ social and political environment is similar in many ways to those of Canada and the United Kingdom, so studies of employment programs in those countries were also eligible if they appeared in the published literature reviews the review team consulted as the first step in identifying potentially eligible studies.

The review team used a two-stage process to determine whether each publication found through the search methods met all eligibility criteria. First, the team examined publications’ titles and abstracts and screened out duplicate citations and those that obviously did not meet the criteria for inclusion. The team performed a second round of screening on the remaining documents, skimming their full text to confirm eligibility and noting the characteristics of the citation and the intervention being examined in a spreadsheet created for this purpose.

III. ASSESSING THE STRENGTH OF EVIDENCE

The review team examined the strength of the evidence of effectiveness presented in the research—that is, how likely it is that a study’s estimated program impacts were *caused by* that program, not something else (an unbiased estimate). This section presents the standards and procedures used to do this. These standards and procedures were designed to ensure that each study’s strength of evidence was accurately characterized for the review’s audiences.

A. Standards

Trained team members examined the research design, applied evidence standards, and assigned a rating to indicate how confident readers can be that the study provides unbiased estimates of the examined intervention’s impact. Reviewers applied the evidence standards to each outcome eligible for review, so studies examining multiple outcomes sometimes received multiple ratings.

1. Eligible study designs

The review assessed two types of eligible designs: randomized controlled trials (RCTs) and comparison group designs.⁴

In RCTs, researchers randomly assign study participants to a treatment group or a control group. The former are allowed to receive the employment program being tested, and the latter are not. RCTs provide the strongest evidence that differences in the observed outcomes between the two groups can be attributed wholly to the program being tested because there are no systematic differences, on average, between groups at the study's onset.

In a comparison group design, researchers typically identify a treatment group that received the services being tested and construct a comparison group that did not receive the services but is otherwise as similar as possible to the treatment group, based on both groups' observed characteristics. The comparison group is intended to show what would have happened to the treatment group had its members not received services. The similarities between the groups on observed characteristics measured at baseline (before the treatment group received any services) help assure that differences in outcomes are the result of the tested program, rather than underlying differences between the group members. However, a comparison group design cannot account for all unobserved factors that might influence both program participation and the outcomes, so it is considered less rigorous than an RCT.

2. Overview of ratings, their interpretation, and relevant criteria

The review had three possible ratings for the strength of evidence on each outcome: high, moderate, or low. As described in Table III.1, the ratings reflect how likely it is that the estimated effects were caused by the program being examined and not by other factors. Only outcomes examined in a study with an RCT design could receive a high rating. Moderate was the highest possible rating for outcomes examined in studies with comparison group designs.

Table III.1. Ratings and their interpretation

Rating	Interpretation
High	There is strong evidence that the effect is solely attributable to the employment program examined.
Moderate	There is moderate evidence that the effect is attributable at least in part to the employment program examined. However, other factors not accounted for in the study might also have contributed to the estimated effect.
Low	There is little evidence that the effect is attributable to the employment program examined. Other factors are likely to have contributed.

The rating for an outcome depended on research design and a combination of other factors discussed in later sections of this report. Overall, a study received the highest rating assigned to any outcome the review examined within the study. An outcome received a **high** rating if it was examined in an RCT in which the combination of overall and differential attrition rates was low

⁴ As indicated earlier, standards for regression discontinuity designs are not discussed in this document because the literature search did not identify eligible studies that used such a design.

and there were no other concerns with the design (such as a problem with the randomization). RCTs with specific problems⁵ and comparison group design studies could have received a **moderate** rating if they showed that the analytic groups are equivalent on baseline characteristics of interest and if the authors controlled for pre-intervention measures of the outcome in the analyses (Figures III.1 and III.2). Outcomes were rated **low** if they are examined in an RCT or comparison group design study that did not receive a high or moderate rating. Specific reasons for this included:

- It was examined in an RCT with specific problems or a comparison group design, *but* the analytic groups were not shown to be equivalent on baseline characteristics of interest.
- It was examined in an RCT with specific problems or a comparison group design, *but* pre-intervention measures of the outcome were not used as controls in the analyses.
- Other factors lined up exactly with the intervention, making it impossible to attribute the estimated effects to the intervention.
- There were systematic differences in data collection between the research groups, calling into question whether estimated impacts were attributable to the program or to the different data collection approaches.

⁵ Including: high attrition, groups not generated randomly, nonrandom allocations of cases after random assignment, or randomly generated groups having different probabilities of assignment over time.

Figure III.1. Flowchart for rating randomized controlled trials

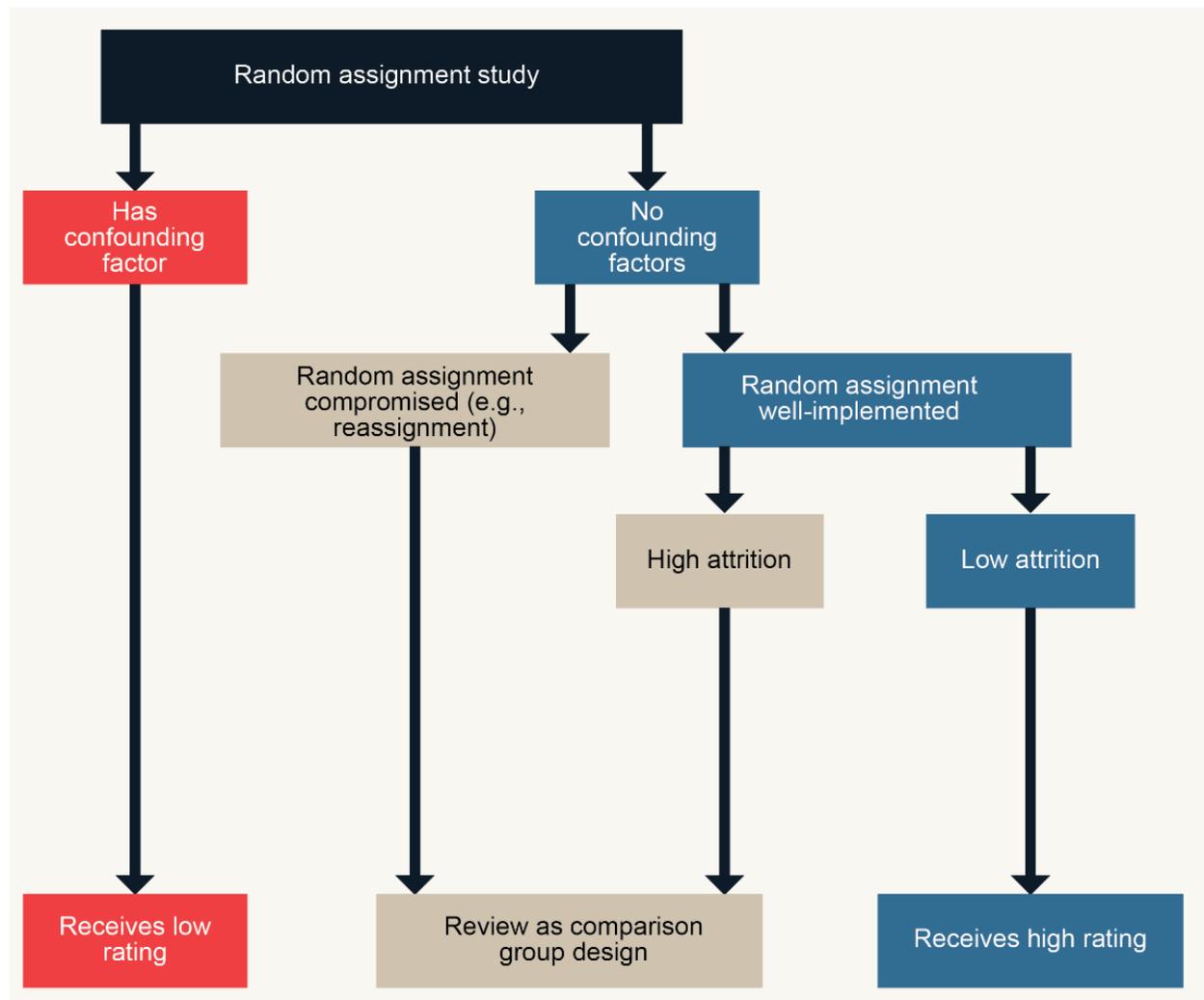
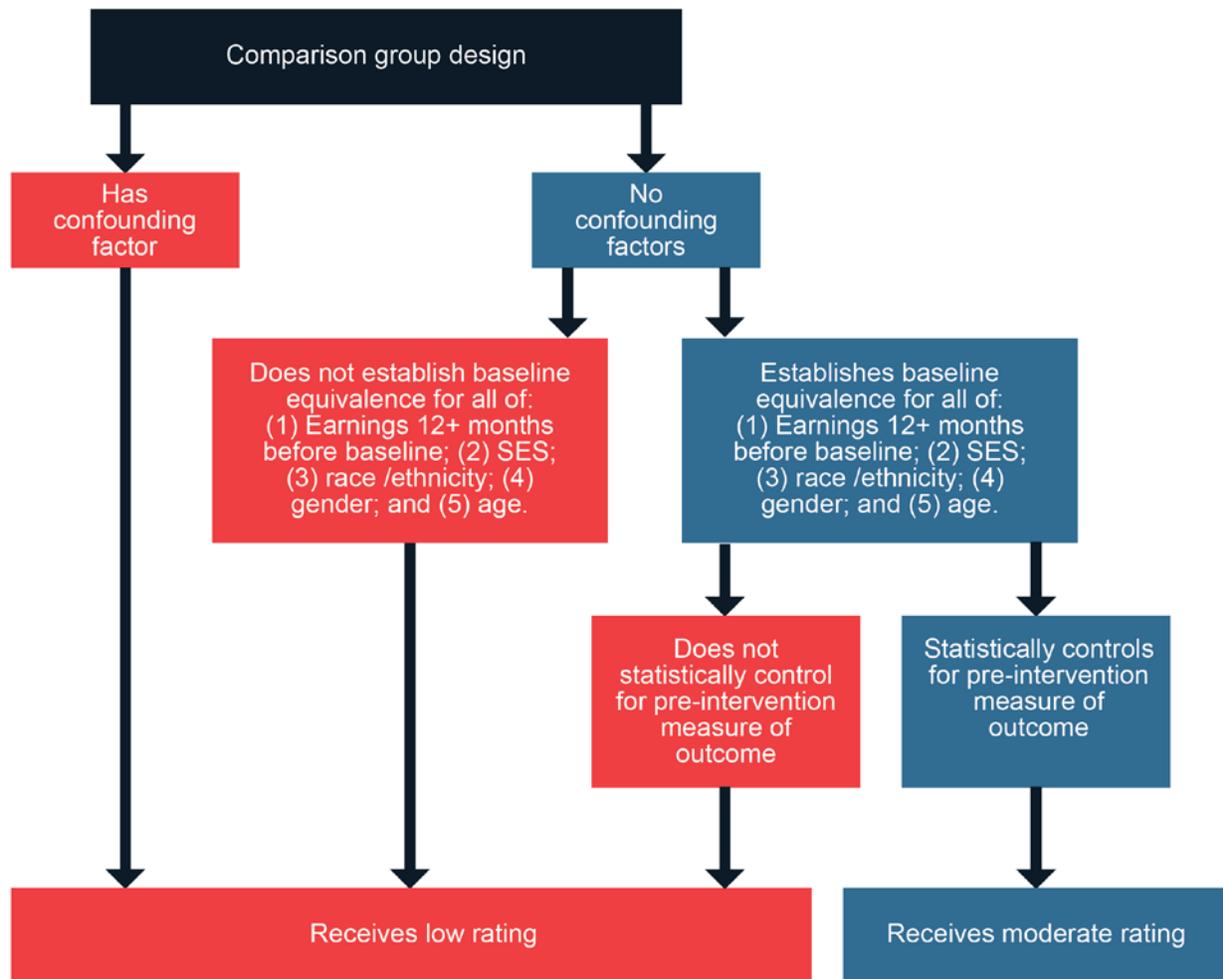


Figure III.2 Flowchart for rating comparison group design studies



3. Assessing randomization procedures

A random assignment design is the strongest possible design because the treatment and control groups are formed by chance, so researchers can be confident that observed differences are due to the program being studied. When reviewers identified deviations from random assignment, such as reassigning or replacing group members or varying the probability of random assignment to each condition over time without adjusting for this variation in the analysis, the review considered such studies as quasi-experimental comparison group designs for the purpose of the review. As such, they were eligible for a rating of moderate at best.

4. Examining attrition in RCTs

Attrition, the proportion of the randomly assigned sample not included in the estimation of effects, is the main determinant of whether estimates from an RCT are free of bias and therefore whether the evidence on the program's effectiveness is strong. Only an RCT with low attrition could receive a high rating. Overall attrition and differential attrition from the treatment and control groups were a concern, because both may lead to bias in the estimated effects. To

determine whether an RCT had low attrition, the project team adopted the attrition boundary used in other systematic reviews sponsored by OPRE. This boundary was developed through an empirical bias model developed by Mathematica Policy Research for the U.S. Department of Education's What Works Clearinghouse.⁶

When the combination of overall and differential attrition rates in an RCT did not exceed the attrition boundaries, the study attrition was considered low and the level of bias acceptable. Table III.2 shows the combinations of overall and differential attrition that were considered low. The review relied on information within each manuscript to calculate attrition. The team occasionally contacted authors for additional information if, for instance, overall attrition was described but information on differential attrition was missing. (If authors did not respond to a request for additional information, reviewers made the conservative assumption that attrition was high.)

Table III.2. 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	1.0
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 3.0.

⁶ The What Works Clearinghouse's conservative boundary is designed to be applied in situations where there is reason for concern that attrition from the study is likely to be associated with outcomes of interest.

RCTs with high attrition were eligible for a moderate rating at most. For RCTs with attrition exceeding the acceptable boundaries, there is substantial risk of bias; therefore, RCTs with high attrition did not receive a high rating. To receive a moderate rating, an RCT with high attrition had to demonstrate baseline equivalence on key characteristics of interest (described in the next section) and had to control for pre-intervention measures of the outcome. High attrition and failure to demonstrate equivalence and/or failure to control for pre-intervention measures of the outcome led to a low rating.

Purposeful selection of a random or time-defined subsample was not considered attrition. Researchers sometimes collected follow-up data from only a subsample of the group initially assigned to treatment and control groups. If this subsample was randomly selected, its non-selected group members were not considered to be lost through attrition.

Imputation does not correct an attrition problem. Researchers may impute missing data using a variety of approaches. Because imputation may mask a high attrition rate, the review assessed attrition based on the unimputed sample.

5. Assessing equivalence on baseline characteristics for RCTs with high attrition and comparison group designs

To receive a moderate rating, RCTs with high attrition⁷ and comparison group designs had to demonstrate that the treatment group and its comparison group had similar observable characteristics at baseline, before the program began. Ensuring that the groups were similar before one group was able to receive services help establish that differences observed between the two groups later were the result of the services.

Although the review adhered closely to the evidence standards used in other OPRE reviews, the review team worked with three researcher consultants to adapt the standards for this topic area. In particular, high-attrition RCTs and comparison group designs had to demonstrate baseline equivalence on the following five characteristics to receive a moderate rating:

1. Earnings, as measured one year or more before baseline⁸
2. Socioeconomic status
3. Race/ethnicity
4. Gender
5. Age

The review considered groups to be equivalent if their means were not statistically different from each other ($p < 0.05$, two-tailed test). If the study authors performed difference-of-means

⁷ In some RCTs with low attrition, authors may have shown that the groups in the analytic samples were not equivalent on baseline characteristics. As long as the random assignment process appeared to have been implemented well and attrition was low, the study did not have to include statistical controls to be eligible for a high rating because the random assignment procedure makes it likely that any differences between the groups at baseline were due to chance alone.

⁸ For employment outcomes, a one-year or more lagged measure of employment was sufficient.

tests, reviewers accepted those results. If authors presented the means and standard deviations for each group but did not test whether the means were statistically different, then the review team performed the tests using the raw data provided.

The review team occasionally requested additional information about equivalence from authors. Sufficient reporting of pre-intervention data had to be included in the study report or obtained from the study authors to allow the review team to draw conclusions about the groups' equivalence. (If authors did not respond to a request for additional information, the review team made the conservative assumption that groups were not equivalent.)

Reviewers considered three conditions of equivalence when assigning a rating:

1. **Was there evidence that the analytic groups were equivalent on socioeconomic status, race/ethnicity, gender, and age at baseline?** If so, and if conditions 2 and 3 were met, the outcomes examined could receive a moderate rating. If equivalence on these demographic factors was not demonstrated for a given analytic sample, this resulted in a low rating for all outcomes examined for that sample, regardless of whether conditions 2 and 3 were met.
2. **Was there evidence that the analytic groups had equivalent earnings at least one year before baseline?** The target groups for employment programs tend to experience unemployment or low wages directly before the program, by definition. But, before their spell of hardship began, these people might have been in different circumstances with respect to their employment and earnings. Therefore, the review was concerned with their experiences a year or more before the study when assessing whether the groups were similar. If there was evidence that the analytic groups had equivalent earnings at least one year before the study began (also known as lagged earnings) and conditions 1 and 3 were met, the outcomes examined could receive a moderate rating. If equivalence on lagged earnings for a given analytic sample was not demonstrated, all analyses that were based on the same analytic sample received a low rating. (However, employment outcomes could still receive a moderate rating if there is evidence that the analytic groups had equivalent lagged employment and conditions 1 and 3 were also met.)
3. **Did the authors control for the baseline or lagged level of the outcome when calculating effects?** Reviewers did not assume that members of each study group in RCTs with high attrition and in comparison group designs were drawn from the same population. Therefore, analyses (even those that showed groups were equivalent on a lagged measure of earnings) had to control for the baseline or lagged level of the outcome to receive a moderate rating (as long as conditions 1 and 2 were also met). Failure to do so resulted in a low rating.

6. Validity of outcome measures

Although 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 they correctly measure the concepts they seek to measure), this review had no such criteria. The review required face validity—that to a general reader, the outcome should seem to measure its intended concept (for example, earnings are not an outcome that has face validity as a measure of educational attainment). Generally, outcomes relevant to the review—such as employment, earnings, and public benefit receipt—tend to be objective and less controversial to measure than

outcomes of primary focus in other reviews (such as co-parenting quality or child behavior). Therefore, there were fewer concerns about those outcomes' validity and reliability.

7. Confounds

Confounds are factors external to the intervention that reach only the members of one study group and could affect outcomes. For instance, if all members of the treatment group received job placement assistance from a single employment counselor, it would be impossible to disentangle the effect of the job placement assistance from the abilities of the employment counselor. The review assigns a low rating to any study in which a reviewer identified a confound.

B. Review process

Each study was assigned for review and was examined by two team members. A first reviewer documented all information pertinent to the evidence rating in a coding guide. A second, highly experienced reviewer or the principal investigator thoroughly checked the review to make sure the evidence rating was correctly applied and the review captured the appropriate information.

Some studies did not contain all the information needed to assign a rating. In these cases, the review requested this information from the study authors. If the authors provided the information, the review team used it to complete the review. If not, the reviewer's rating was based on the information in the study. This meant, for instance, that comparison group designs lacking information on baseline equivalence received a low rating if the authors did not respond to a request for further information.

C. Conflicts of interest

All reviewers were required to sign a certification disclosing any potential conflicts of interest, such as having participated in conducting the study being reviewed. Individuals were not permitted to review studies if they have a conflict of interest. Their signed statements are filed with OPRE. All eligible studies conducted by Mathematica were reviewed by an individual who does not work at Mathematica. A senior reviewer from Mathematica checked that review, but all final decisions on the study and outcome ratings had to be affirmed by the external reviewer.

IV. EXTRACTING AND PRESENTING DATA

Reviewers systematically recorded information about a study, and about selected outcomes within the study, in a template. The templates were combined into a database of results that is searchable on the review's website. If a single publication reported impacts for more than two distinct contrasts (such as multiple programs or sites), the review reported results separately for each pair of contrasts and provided links to the related results.

Using the review template, reviewers captured standard information about each study and outcome, as described in Table IV.1. The review did not publish results from studies earning a low rating.

Table IV.1. Information captured at the study and outcome level during study reviews

Study-level data	Outcome-level data
Full and brief citation	Findings for preferred outcomes by data source and domain
Name of evaluation	Means for treatment and comparison groups
Name of the program studied	Impact
Characteristics of study participants	Indication of whether the impact is favorable, unfavorable, or there is no statistically significant ($p < 0.05$) effect
Setting	Data sources and timing of data collection
Design (how researchers formed treatment and comparison groups)	Sample size
Treatment condition and duration	Rating of the strength of the evidence
Comparison condition	
Study funding	
Program history and funding	
Timing of study	
Study rating and reason for any rating other than high	
<ul style="list-style-type: none"> • Method for calculating effect • Brief description of subgroups examined by the study (although subgroup results are usually not presented) • Outcome domains examined^a 	
URL to publicly available full text, when possible	

^a Reviewers noted whether the study reported findings in certain secondary outcome domains (physical health, mental health, substance abuse, criminal justice, housing, nutrition, financial assets, parenting and co-parenting, couple relationships, family formation, child well-being), but the review did not report the findings in these domains.

The review focused on outcomes in the following domains: short- and long-term employment, short- and long-term earnings, short- and long-term public benefit receipt, and acquisition of education and/or training. Short-term was defined as 18 months or less.

Although study authors may have had certain outcomes in mind when they begin analysis (for instance, annual earnings and employment rate), they may ultimately have examined more finely grained detail to identify effects. This type of exploration can lead authors to report dozens of outcomes that all essentially measure the same underlying construct. Thus, to assure that the review results are focused and manageable to the user, the review team developed a set of decision rules for selecting preferred specifications or levels of detail for key outcomes in each domain. To ensure that the review preserved at least some information from each data source, reviewers applied these rules *within each domain, to each data source*. This means that reviewers identified all the data sources analyzed in a study, determined which domains the outcomes from those data sources examined, and applied the rules to focus on preferred outcomes within each data source-domain combination. Each survey administration was treated as a separate data source, even if the survey periods were both long-term or both short-term. The approach of pre-defining criteria for selecting among multiple outcome measures is endorsed by

the Campbell Collaboration as a highly desirable method for reporting results in systematic reviews.⁹

The review always preferred analyses of the full sample. However, in some instances data were only available for a subgroup of the total study population; for example, a survey fielded to a randomly selected subsample of participants, or a cohort enrolled at a specific time, or a subgroup of study sites with administrative records available. The review reported subgroup results for a study only when no full group results were available for a given data source in a given domain.

Table IV.2 displays the hierarchy of preferred outcomes for reporting, and indicates the review's preferred outcomes in bold text.

V. CONCLUSION

The review of employment related programs for low-income adults presents practitioners, policymakers, researchers, and the general public with a transparent and systematic assessment of research evidence. The review team systematically identified the body of most highly relevant effectiveness studies by combing through published and unpublished literature using key terms and screening criteria. Then, trained reviewers assessed whether each study had a sufficiently strong design (rating high or moderate) to support confidence in the reported findings of the studies. Finally, the team extracted selected, highly relevant outcomes in several areas related to employment, training, and self-sufficiency, to report on a public website. The website allows users to search for this information and apply filters to identify employment-related programs and strategies of greatest interest to them.

⁹ Methodological Expectations of Campbell Collaboration Intervention Reviews, version 1.0. http://www.campbellcollaboration.org/artman2/uploads/1/MEC2IR_conduct_standards_v1_0_Updated_September_2014.docx. Accessed November 12, 2014.

Table IV.2. Decision rules for selecting outcomes to report within each domain for each data source

Domain(s)	Rules
Employment (short- and long-term)	<ul style="list-style-type: none"> ✓ Include employed at the time of follow-up (or each follow-up, if applicable) in any job. If a subsidized job, ideally report both overall and unsubsidized job outcomes, otherwise report both subsidized and unsubsidized. For administrative data, this may be the most recent month or quarter. <ul style="list-style-type: none"> ✓ If not available, use “ever employed” (for example, since random assignment/program completion). ✓ If only reported by full-time/part-time status, and not an overall employment rate, use those two. ✓ Include cumulative measures of employment, if available (for example, percent employed more than one quarter, percent employed more than two quarters, number of consecutive quarters employed), <ul style="list-style-type: none"> ✓ But, do not include other decompositions of employment (for example, employed in quarter 1, employed in quarter 2). X Do not include employment by job characteristics (for example, percent employed in a job paying more than \$11/hour, percent employed in a job offering benefits).
Earnings (short- and long-term)	<ul style="list-style-type: none"> ✓ Include annual earnings for the longest elapsed follow-up year (for example, earnings in year 3 of a 3-year follow-up period) in any job. If a subsidized job, ideally report both overall and unsubsidized job outcomes, otherwise report both subsidized and unsubsidized. <ul style="list-style-type: none"> ✓ If not available, use <u>average annual</u> earnings over the follow-up period (for example, average annual earnings years 1-8). ✓ If not available, use <u>aggregate</u> earnings over the entire follow-up period (for example, total earnings since random assignment). ✓ If not available, use <u>quarterly (ideally) or monthly (if quarterly is unavailable) earnings for the longest follow-up period</u> (for example, earnings in quarter 16 or month 48 of a 4-year follow-up period). ✓ If no earnings are presented, use average/median hourly wage as reported at the time of the follow-up. X Do not include earnings outcomes if they are calculated just among those who were employed.
Benefit receipt (short- and long-term)	<ul style="list-style-type: none"> ✓ Include 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). (Specific benefit programs are of policy interest to OPRE. Therefore, the review will report the results for decompositions of benefit receipt if they are presented by study authors for the full analytic sample (for example, measures of receiving TANF, SNAP, UI). <ul style="list-style-type: none"> ✓ If not available, use 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). ✓ If not available, use 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). ✓ If amounts listed above are not available, use quarterly (ideally) or monthly (if quarterly is unavailable) benefit receipt for the longest follow-up period (for example, benefits collected in quarter 16 or month 48 of a 4-year follow-up period). X Do not include amount of benefits if it is reported only for those who received benefits.
Education/training	<ul style="list-style-type: none"> ✓ Include measures of educational attainment over the follow-up period (for example, acquisition of a GED, associate’s degree, BA, attaining a certificate/credential). X Do not include decompositions of these measures (for example, attained a GED within 1 year, attained a GED within 2 years). X Do not include outcomes that combine these individual educational attainment indicators in different ways (for example, attained a GED or completed training, attained an associate’s degree or credential), unless the individual measures are not presented.

Note: The review reports at least **one outcome from each domain by data source combination** examined by the authors.

APPENDIX A

LITERATURE REVIEWS

- Bloom, Dan. “Transitional Jobs: Background, Program Models, and Evaluation Evidence.” New York: MDRC, February 2010.
- Bloom, Dan, Pamela J. Loprest, and Sheila R. Zedlewski. “TANF Recipients with Barriers to Employment.” Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation, August 2011.
- Decker, Paul T., and Jillian A. Berk. “Ten Years of the Workforce Investment Act (WIA): Interpreting the Research on WIA and Related Programs.” *Journal of Policy Analysis and Management*, vol. 30, no. 4, 2011, pp. 906–926.
- Greenberg, David, Charles Michalopoulos, and Philip Robins. “A Meta-Analysis of Government-Sponsored Training Programs.” *Industrial and Labor Relations Review*, vol. 57, no. 1, October 2003, pp. 31–53.
- Greenberg, David, Philip Robins, and Robert Walker. “Conducting Meta-Analyses of Evaluations of Government-Funded Training Programs.” *Review of Policy Research*, vol. 22, no. 3, 2005, pp. 345–367.
- Greenberg, David, Victoria Deitch, and Gayle Hamilton. “Welfare-to-Work Program Benefits and Costs: A Synthesis of Research.” New York: MDRC, 2009.
- Greenstone, Michael, and Adam Looney. “Building America’s Job Skills with Effective Workforce Programs: A Training Strategy to Raise Wages and Increase Work Opportunities.” Washington, DC: Brookings Institution, November 2011.
- Hamilton, Gayle. “Temporary Assistance for Needy Families Program—Research Synthesis Brief Series.” Research Brief #06: Improving Employment and Earnings for TANF Recipients. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation, March 2012.
- Hamilton, Gayle, and Susan Scrivener. “Increasing Employment Stability and Earnings for Low-Wage Workers: Lessons from the Employment Retention and Advancement (ERA) Project.” OPRE Report 2012-19. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation, April 2012.
- Klerman, Jacob, Robin Koralek, Ashley Miller, and Katherine Wen. “Job Search Assistance Programs—A Review of the Literature.” OPRE Report 2012-39. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation, December 2012.
- Mastri, Annalisa, Emily Evans, Stacy Pancratz, and Carrie Wolfson. “Employment Research in Brief: An Annotated Bibliography of ETA-Sponsored Studies.” Princeton, NJ: Mathematica Policy Research, November 2011.
- Michalopoulos, Charles. “What Works Best for Whom: Effects of Welfare and Work Policies by Subgroup.” New York: MDRC, March 2004.

What Works Clearinghouse. “Procedures and Standards Handbook: Version 3.0.” Washington, DC: Institute of Education Sciences, March 2014. Available at: http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v3_0_standards_handbook.pdf. Accessed October 17, 2014.

APPENDIX B

CALL FOR PAPERS

Call for Papers: Effectiveness Studies of Employment Strategies for Low-Income Adults

SUBMISSION DEADLINE: July 17, 2014

Mathematica Policy Research is seeking completed research papers or articles examining the effectiveness of programs or strategies whose primary aim is to improve employment outcomes for low-income adults.

This Call for Papers is part of an effort to comprehensively identify research for the Employment Strategies for Low-Income Adults Evidence Review, which Mathematica is conducting for the Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

We are particularly interested in unpublished manuscripts (past or recent), conference papers, and forthcoming articles that are not included in existing research syntheses or databases. Please include papers even if available online—for example, those on university or program web sites.

Papers submitted in response to this call should:

- Use a quantitative approach to examine the effectiveness of a program or strategy whose primary aim is to improve the employment-related outcomes of low-income adults.
 - Studies may examine programs or strategies. Programs may use one or more strategies. Examples of strategies include but are not limited to: job search assistance, work experience, job training, personal supports, and financial incentives.
 - Examples of employment-related outcomes include but are not limited to: becoming employed, being employed, maintaining employment, occupational advancement, employment with benefits, income, earnings, and public assistance receipt.
- Examine a strategy or program serving individuals aged 18 and older. Programs may also serve individuals younger than age 18, but must include individuals 18 and older as well.
- Have been prepared or published in 1990 or later.
- Examine a strategy or program conducted in the United States of America.

If studies of strategy or program implementation are available that accompany the effectiveness study, call respondents are invited to submit those as well. Standalone implementation studies of a program for which no effectiveness study is available will not be included in the review.

The purpose of this review is to provide practitioners, policymakers, researchers, and the general public with a transparent and systematic assessment of the effectiveness of approaches for improving the employment-related outcomes of low-income individuals. Reviewers will assess the strength of the evidence of effectiveness provided by individual studies and share results on a publicly available website with a searchable database of findings.

Please feel free to forward this call to other researchers, practitioners, or organizations that may have relevant studies.

SUBMISSION INSTRUCTIONS

Submissions should be in MS Word or PDF format. Files may be sent to:
employmentreview@mathematica-mpr.com

The deadline for submissions is July 17, 2014.

Respondents will be notified that we received their submission, but the notice will not indicate whether the paper might be included in the review.

www.mathematica-mpr.com

Improving public well-being by conducting high quality,
objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ WASHINGTON, DC

MATHEMATICA
Policy Research

Mathematica® is a registered trademark
of Mathematica Policy Research, Inc.