

# A GUIDE FOR USING ADMINISTRATIVE DATA TO EXAMINE LONG-TERM OUTCOMES IN PROGRAM EVALUATION

OPRE Report  
2021-145

August 2021

**Executive Summary**



# A Guide for Using Administrative Data to Examine Long-Term Outcomes in Program Evaluation

OPRE Report 2021-145

August 2021

AUTHORS: Jonathan Bigelow, Alexandra Pennington,  
Kelsey Schaberg, and Deondre' Jones, MDRC

SUBMITTED TO:  
Brett Brown, Project Officer  
Office of Planning, Research, and Evaluation, Administration for Children and  
Families, U.S. Department of Health and Human Services

PROJECT DIRECTOR:  
Kelsey Schaberg  
MDRC  
200 Vesey Street  
New York, NY 10281

CONTRACT NUMBER: HHSP233201500059I

This report is in the public domain. Permission to reproduce is not necessary.

SUGGESTED CITATION: Bigelow, Jonathan, Alexandra Pennington, Kelsey Schaberg, and Deondre' Jones 2021. *A Guide for Using Administrative Data to Examine Long-Term Outcomes in Program Evaluation*. OPRE Report 2021-145. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

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 [www.acf.hhs.gov/opre](http://www.acf.hhs.gov/opre).



[Sign-up for the OPRE Newsletter](#)



Follow OPRE on  
Twitter [@OPRE\\_ACF](https://twitter.com/OPRE_ACF)



Like OPRE on Facebook  
[facebook.com/OPRE.ACF](https://facebook.com/OPRE.ACF)



Follow OPRE on Instagram  
[@opre\\_acf](https://www.instagram.com/opre_acf)



# Funding and Acknowledgments

This guide is funded by the U.S. Department of Health and Human Services, Administration for Children and Families (ACF), Office of Planning, Research, and Evaluation (OPRE), as part of the “From Theory to Practice” (T2P) project.

First, we thank Brett Brown, the T2P project officer from OPRE, who helped sculpt the overall vision for the guide and provided valuable input on working drafts. We would also like to thank Bethanne Barnes and Aaron Goldstein from OPRE for their thoughtful suggestions.

The guide also benefited from the guidance offered by members of the T2P project’s expert panel: Daniela Hochfellner from the Center of Urban Science and Progress at New York University, Hilary Hoynes from UC Berkeley, Lynn Overmann from Opportunity Insights, and Lisa Morel from the National Center for Health Statistics (NCHS) of the Centers for Disease Control and Prevention (CDC).

The authors would also like to thank several MDRC staff members who provided support and guidance on the production of the guide. Shawna Anderson oversaw and managed the budget. Richard Hendra and Mark van Dok provided thoughtful input and feedback on several drafts of the guide. Alice Tufel offered guidance on the structure and content of the guide. Rebecca Bender edited the guide and Carolyn Thomas prepared it for publication.

Dissemination of MDRC publications is supported by the following organizations and individuals that help finance MDRC’s public policy outreach and expanding efforts to communicate the results and implications of our work to policymakers, practitioners, and others: The Annie E. Casey Foundation, Arnold Ventures, Charles and Lynn Schusterman Family Foundation, The Edna McConnell Clark Foundation, Ford Foundation, The George Gund Foundation, Daniel and Corinne Goldman, The Harry and Jeanette Weinberg Foundation, Inc., The JPB Foundation, The Joyce Foundation, The Kresge Foundation, and Sandler Foundation.

In addition, earnings from the MDRC Endowment help sustain our dissemination efforts. Contributors to the MDRC Endowment include Alcoa Foundation, The Ambrose Monell Foundation, Anheuser-Busch Foundation, Bristol-Myers Squibb Foundation, Charles Stewart Mott Foundation, Ford Foundation, The George Gund Foundation, The Grable Foundation, The Lizabeth and Frank Newman Charitable Foundation, The New York Times Company Foundation, Jan Nicholson, Paul H. O’Neill Charitable Foundation, John S. Reed, Sandler Foundation, and The Stupski Family Fund, as well as other individual contributors.

For information about MDRC and copies of our publications, see our website: [www.mdrc.org](http://www.mdrc.org).

# Overview

## INTRODUCTION

Many social programs are intended to generate long-term benefits for their participants, but evaluations of those programs have historically not had access to the necessary resources to measure such outcomes over the long run—for 5, 10, or 15 years or longer. Administrative data—data that are created and stored to enable government administration, or as a by-product of it—present a potentially low-cost opportunity for tracking the long-term effects of new policy or program interventions. However, the procedures for gaining access to these data are often idiosyncratic, time-intensive, or undocumented.

The case of the Moving to Opportunity demonstration may be particularly instructive here. Early research focused on the adults of households that were supported in moving from subsidized, public housing to neighborhoods with low levels of poverty, finding little to no economic impact after families completed such moves. However, later findings indicated that living in neighborhoods with low poverty levels had substantial, positive economic impacts on some children of those families after they reached adulthood. As government agencies and their research partners consider opportunities to leverage these data to extend evidence about their programs—and as data privacy and security take on ever-increasing importance—the Administration for Children and Families’ Office of Planning, Research and Evaluation (ACF/OPRE) in the U.S. Department of Health and Human Services is developing resources to support these interests and explore the benefits and limitations of linking study and administrative data for long-term research. The “From Theory to Practice” project represents one of ACF/OPRE’s efforts to support the research community in conducting such explorations.

## PURPOSE

This *Guide for Using Administrative Data to Examine Long-Term Outcomes in Program Evaluation* is being produced to complement federal efforts to expand the use of administrative data for building evidence—in this case, evidence about the long-term effectiveness of federally funded programs and interventions. This guide is a resource to assist program evaluation teams—including funders, sponsors, and evaluation research partners—in assessing the feasibility and potential value of examining long-term outcomes using administrative data. It describes common steps that are involved in linking evaluation data and administrative data. It will help teams tackle topics such as:

- how to identify worthwhile, policy-relevant opportunities for extending evaluation follow-up
- what study data and infrastructure are required to enable extended follow-up

- factors to consider in selecting suitable administrative data sources
- navigating the legal and ethical requirements that are commonly associated with pursuing extended follow-up
- special considerations for matching study and administrative data
- how to assess the quality of linked study and administrative data

The guide is directed primarily toward research teams considering whether to examine long-term outcomes for evaluations, particularly those whose work has been completed. This guide also includes valuable information that may enable research on long-term outcomes for new or ongoing evaluations.

## KEY FINDINGS AND HIGHLIGHTS

This guide proposes to think about the preparation and work that are required to extend study follow-up using administrative data in three main phases of effort:

- Phase 1 entails considering the value and practicality of long-term follow-up. It is focused on ensuring that there is a solid policy and research justification for extending follow-up and that there are suitable and accessible administrative data that will answer specified research questions.
- Phase 2 involves preparing for long-term follow-up by laying the necessary legal and ethical groundwork. Notably, study teams are advised to ensure that data-related agreements between evaluation teams and other entities governing the research describe and enable the planned research activities. Teams are also advised to assess what, if any, human subjects ethical standards apply to the proposed long-term follow-up research by consulting an Institutional Review Board (IRB).
- Phase 3 centers around assessing administrative data to determine whether they are suitable for answering the proposed research questions and linking to study data. Researchers are advised to assess administrative data providers' requirements for matching to study data, to consider the extent to which administrative data will cover study participants and their activities, to determine the identifiers that will be used to match and the method for matching, and to establish how the quality of linked data will be assessed.

The full guide describes in more detail the various considerations that study teams might take into account to begin to realize the potential uses of administrative data in researching long-term outcomes. Examples and case studies throughout the guide generally highlight efforts to research long-

term economic outcomes, such as participant employment and earnings, but the concepts presented should be applicable to a variety of social policy research contexts.

## **GLOSSARY**

**Administrative data:** Information created and stored to enable government administration, or as a by-product of it. In this context, administrative data may allow researchers to describe and analyze the experiences and outcomes of particular interest to evaluations of federally-funded social programs.

**Follow-up:** A period of time across which a program evaluation expects to describe individual study participants' activities and outcomes.

**Identifier:** A set of numbers or characters, such as a Social Security number or name, that can be used to identify an individual, either on its own or in combination with other identifiers.

**Institutional Review Board (IRB):** A type of formally designated committee that applies research ethics standards and statutes. IRBs review the methods that are proposed for research to ensure that they should be conducted and that they take appropriate steps to protect the rights and welfare of humans participating as subjects in a research study.

**Long-term outcomes:** The findings among individuals participating in a social program, as measured in a program evaluation over a period of time and generally considered to be longer than 5 or 10 years.

# Executive Summary

Many social programs are intended to generate long-term benefits for their participants, but evaluations of those programs have historically not had access to the necessary resources to measure those outcomes over the long run—for 5, 10, or 15 years or longer. The case of the Moving to Opportunity demonstration may be particularly instructive here. Early research focused on the adults of households that were supported in moving from subsidized, public housing to neighborhoods with low levels of poverty, finding little to no economic impact after families completed such moves. However, later findings indicated that living in neighborhoods with low poverty levels had substantial, positive economic impacts on some children of those families after they reached adulthood.<sup>1</sup> Administrative data—data that are created and stored to enable government administration, or as a by-product of it—present a potentially low-cost opportunity for tracking the long-term effects of new policy or program interventions. However, the procedures for gaining access to these data are often idiosyncratic or time-intensive. In addition, little documentation is available about how to access and use these data for research purposes, and researchers are likely to encounter unique data-quality challenges in so doing.

This *Guide for Using Administrative Data to Examine Long-Term Outcomes in Program Evaluation* is being produced as part of the “From Theory to Practice” (T2P) project. This guide complements federal efforts to expand the use of administrative data for building evidence—in this case, evidence about the long-term effectiveness of federally funded programs and interventions. Through T2P, the Administration for Children and Families’ Office of Planning, Research, and Evaluation (ACF/OPRE) in the U.S. Department of Health and Human Services is developing resources to support these interests and explore the benefits and limitations of linking study and administrative data for long-term research.

This guide is a resource to assist program evaluation project teams—including funders, sponsors, and evaluation research partners—in assessing the feasibility and potential value of examining long-term outcomes using administrative data. It describes common steps that are involved in linking evaluation data and administrative data. The guide will help teams tackle topics such as:

- how to identify worthwhile, policy-relevant opportunities for extending evaluation follow-up
- what study data and infrastructure are required to enable extended follow-up

---

<sup>1</sup>Raj Chetty, Nathaniel Hendren, and Lawrence F. Katz, “The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Project,” *American Economic Review* 106, 4 (2016): 855–902.

- factors to consider in selecting suitable administrative data sources
- navigating the legal and ethical requirements that are commonly associated with pursuing extended follow-up
- special considerations for matching study and administrative data
- how to assess the quality of linked study and administrative data

The guide is directed primarily toward research teams considering whether to examine long-term outcomes for evaluations, particularly those whose work has been completed. This guide also includes valuable information that may enable research on long-term outcomes for new or ongoing evaluations. Examples and case studies throughout the guide generally highlight efforts to research long-term, employment-related interventions, but the concepts presented should be applicable to a variety of social interventions.

## A SUMMARY OF CONCEPTS INCLUDED IN THIS GUIDE

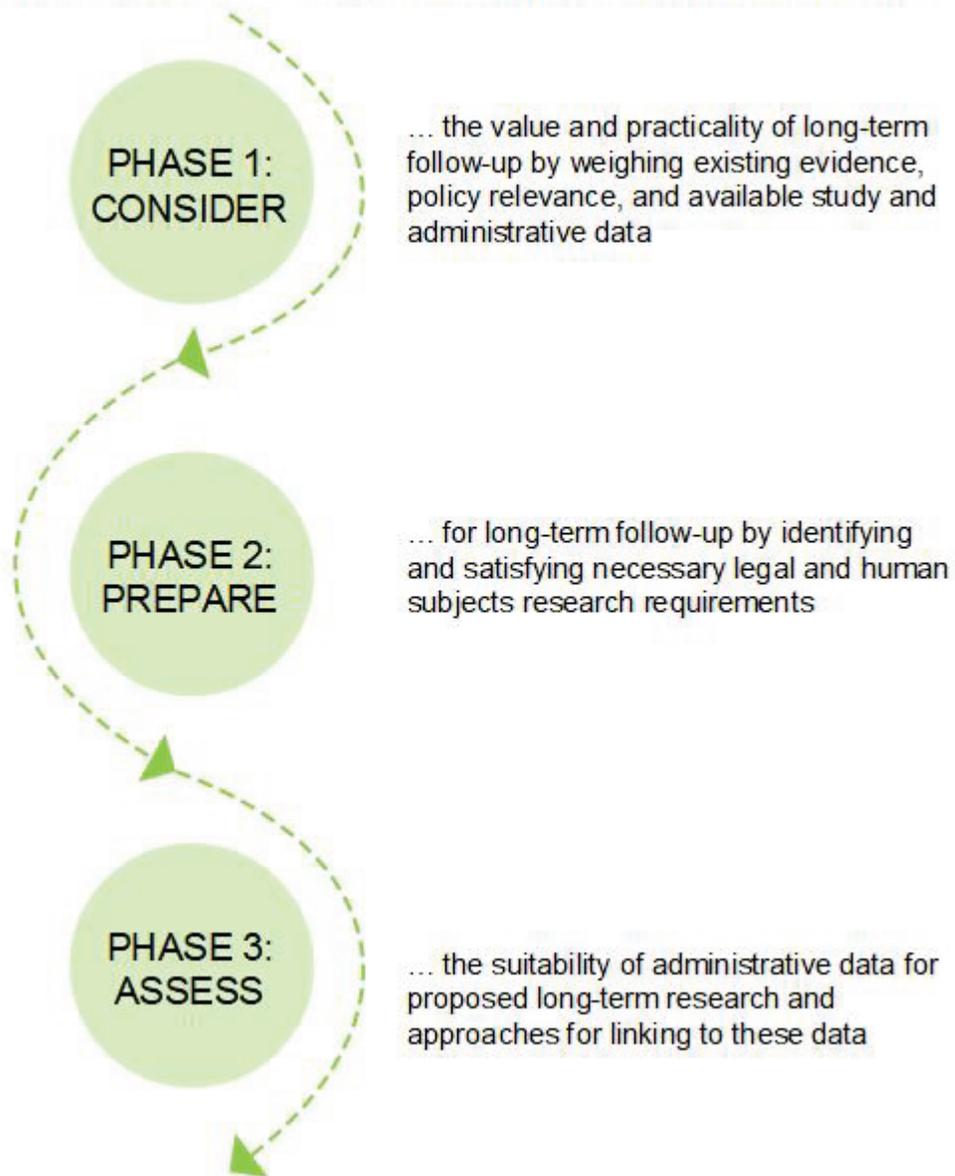
This guide proposes to think about the preparation and work that are required to extend study follow-up using administrative data in *three main phases of effort*. (See Figure ES.1.)

**Phase 1** entails *considering the value and practicality of long-term follow-up*, ensuring that there is solid policy and research justification to extending follow-up and that there are suitable and accessible administrative data that will answer specified research questions. Study teams are advised to:

- Articulate a theory of change and analysis plan, grounded in existing evidence, that prospectively outlines and justifies the proposed research on long-term outcomes.
- Ensure that proposed long-term follow-up will yield credible, unbiased results about long-term effects and that no threats to internal validity from prior research pose risks to proposed new activities.
- Assess whether the proposed analyses will have adequate statistical power to detect meaningful effects.
- Inventory extant study data to confirm that necessary data, especially participants' personally identifiable information (PII), are available for linking to administrative records.
- Consider the possible sources of administrative data to measure outcomes of interest and identify factors that should be taken into account during planning, including what data are available, the

**Figure ES.1**

**Three Phases of Work for Extending Follow-Up Using Administrative Data**



process for obtaining them, how matching between data sets will occur, and whether there are significant costs or restrictions associated with accessing them.

Tackling this work will vary by the circumstances and aspirations of each study and proposed analysis. However, there are common tasks and conditions shared by studies based on the stage of work that each is in.

- Studies in the planning stage have the chance to work prospectively to lay the necessary groundwork for future research on long-term outcomes even before participants are enrolled and initially exposed to the interventions that are being evaluated. Such studies should generate program theory hypothesizing long-term effects, invest in writing informed consent forms and procedures that enable extended follow-up, and include language in data agreements that will smooth the path toward future long-term follow-up.
- Ongoing studies may (re)assess the prospects for long-term follow-up given what is known about the implementation of studied interventions and their contexts, and consider any course corrections in proposed long-term research plans to account for new learnings about implementation or developments in the data landscape.
- Finally, studies whose primary or initial work has been completed should inventory study data artifacts, such as funding data agreements, to assess feasibility and inform planning for research investigating outcomes that may (or may not) differ significantly from those that were examined during earlier efforts.

**Phase 2** involves *preparing for long-term follow-up by laying the necessary legal and ethical groundwork*. After research teams have assessed the value and practicality of conducting long-term follow-up research, there are two essential hurdles to clear.

- Study teams must ensure that data-related agreements that govern the administration and scope of the study—including funding agreements, agreements with site partners, agreements with participants (such as informed consent documents), and agreements with data providers—adequately describe and enable the planned research activities.
- Researchers must assess what, if any, human subjects ethical standards apply to the proposed long-term follow-up research by consulting an Institutional Review Board (IRB). In particular, teams must establish whether the research is subject to the Common Rule. The Common Rule encompasses the baseline ethical standards under which government-funded (and many other types of) research in the United States is conducted. At least one IRB with jurisdiction will, in most cases, review proposed long-term follow-up activities. Determining whether there is a requirement to collect new informed consent will likely be central to an IRB’s review of proposed new research activities. It will also consider factors such as the burdens and benefits of the research and the data privacy and security measures to be employed by researchers.

Studies in the planning stage are more likely than not to have the ability to anticipate future long-term research activities by describing them in data-related agreements, IRB applications, and informed consent forms, in consultation with funders, data providers, IRB(s), and site and other partners. Studies that are still ongoing or that are completed will probably have to reconcile plans for research on long-term follow-up with past agreements and documents, amending agreements as

necessary, and considering whether participants can (or must) be reconsented, or if waivers of consent may be secured.

After necessary clearances and permissions are obtained, **Phase 3** centers around *assessing administrative data to determine if they are suitable for answering the proposed research questions*. Considerations that study teams must confront often include the following:

- What an administrative data provider's requirements are for how study data and administrative data can be matched. The matching process may be as straightforward as researchers sharing a sample file with study identifiers (for example, names and Social Security numbers [SSNs]), with providers then returning a file with additional measures merged onto those records. Some data providers use more elaborate procedures that protect individuals by returning anonymized data to the research team, and research teams may consider approaches to ensuring that the ability to conduct analyses using administrative data is not constrained.
- The extent to which administrative data cover study participants and the desired activity and time period that are being investigated. Determining what match rates between study data and administrative data are reasonable to expect may be more art than science, depending on the study population, the program context, and the nature of the data source in question.
- The person-level identifiers that will be used for matching, and what type of matching method will be used to link data. In many cases, exact matching on identifiers such as SSNs will be possible, while other data providers may use probabilistic matching methods, using a combination of fields that are unlikely to change (for example, race, gender) to identify records with a high probability of the matches being "true matches."
- How the quality of the linked data can be assessed. Assessing match rates overall and by research groups (for example, for the program and control groups) is customary when assessing data matches, as is investigating the characteristics of study sample members who did not match to administrative records for any systematic trends (for instance, a lower match rate from a certain study site or demographic group).

Long-term matches can present unique considerations for study teams, including:

- certain identifiers, such as case numbers, that are less reliable over time
- participant geographic mobility, such as moves across state lines in studies that expect to rely on data sources from the state included in the original study only
- the use of different data sources for long-term research than those that were used for earlier research

- secular changes to social programs, such as time limits or program rules that might change the extent to which participants are expected to receive those benefits in the long run and therefore be covered by data sources capturing those benefits
- changes to administrative data coverage, quality, or the process for accessing them, such as new laws dictating the extent to which researchers may access data

The full guide describes in more detail the considerations that should be taken into account by study teams aspiring to research the long-term outcomes of social programs, offering examples, case studies, and links to additional resources.