OPRE Report #2015-97 • October 2015

Advancing Evidence-Based Decision Making: **A Toolkit on Recognizing and Conducting Opportunistic Experiments in the Family Self-Sufficiency and Stability Policy Area**

Advancing Evidence-Based Decision Making: **A Toolkit on Recognizing and Conducting Opportunistic Experiments in the Family Self-Sufficiency and Stability Policy Area**

OPRE Report #2015-97 • October 2015

Alicia Meckstroth • Alexandra Resch • Jonathan McCay • Michelle Derr • Jillian Berk • Lauren Akers

Submitted to:
Office of Planning, Research, and Evaluation
Administration for Children and Families
U.S. Department of Health and Human Services
Project Officer: Emily Schmitt

Contract/Task Number: HHSP23320095642WC / HHSP23337043T

Submitted by:
Mathematica Policy Research
1100 1st Street, N.E. Suite 1200
Washington, DC 20002-4221
Project Director: Michelle Derr

This report is in the public domain. Permission to reproduce is not necessary. Suggested citation: Meckstroth, Alicia, Alexandra Resch, Jonathan McCay, Michelle Derr, Jillian Berk, and Lauren Akers (2015). *Advancing Evidence-Based Decision Making: A Toolkit on Recognizing and Conducting Opportunistic Experiments in the Family Self-Sufficiency and Stability Policy Area*, OPRE Report #2015-97, 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.

OPRE

MATHEMATICA Policy Research

Acknowledgements

This report was developed by Mathematica Policy Research, under contract to the U.S. Department of Health and Human Services (HHS), Administration for Children and Families (ACF), Office of Planning, Research, and Evaluation (OPRE). The report was prepared as part of the OPRE-funded project Advancing Welfare and Family Self-Sufficiency Research (Project AWESOME).

Special thanks to several contributors are in order. First, the authors are grateful for the guidance and support of OPRE's Brendan Kelly and Emily Schmitt during the planning and preparation phases of this report. Second, many thanks to Mary Farrell and Mike Fishman from MEF Associates for their contributions related to the case studies. Third, the report benefitted and was adapted from a recently published report funded by the U.S. Department of Education: Resch, A., J. Berk, and L. Akers. *Recognizing and Conducting Opportunistic Experiments in Education: A Guide for Policymakers and Researchers* (REL 2014–037). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Analytic Technical Assistance and Development, 2014.

The authors gratefully acknowledge these contributions and accept sole responsibility for any errors or omissions in this report. The views expressed are those of the authors and do not necessarily represent the views or policies of HHS/ACF/OPRE.

Contents

Acknowledgements ... 3

I: Making the Case for Opportunistic Experiments ... **5**

How and in what settings are opportunistic experiments used? ... 7

What are the defining features of opportunistic experiments? ... 8

What program situations provide opportunities for experimentation? ... 10

In conclusion: making the case for opportunistic experiments ... 10

II: Recognizing Potential Opportunistic Experiments and Assessing Their Feasability ... **14**

Step 1. Look for potential random assignment opportunities ... 14

Step 2. Consider how opportunities align with program and research needs ... 15

Step 3. Assess the feasibility of the research project ... 16

III: Conducting Opportunistic Experiments ... **25**

Step 1. Identify sites ... 25

Step 2. Encourage participation ... 25

Step 3. Conduct random assignment and monitor compliance ... 26

Step 4. Collect data ... 28

Step 5. Analyze data and report findings ... 28

In conclusion ... 28

References ... **30**

TOOLKIT MODULES

Module 1 defines and describes opportunistic experiments (OE) and highlights several case studies.

Module 2 outlines key steps in recognizing potential OEs and assessing their feasibility.

Module 3 provides an overview of the key steps in conducting an OE.

TOOLKIT MODULE 1

About this research guide: This document is the first in a set of three modules comprising the full "Toolkit on Recog nizing and Conducting Opportunistic Experi ments in the Family Self-Sufficiency and Stability Policy Area."

I. Making the Case for Opportunistic Experiments

Randomized controlled trials (RCTs) are considered the gold standard for evaluating an intervention's effectiveness. RCTs provide the rigorous evidence of efficacy that is essential for policymakers and administrators to make evidence-based policy and program decisions. As human services agencies continue to face limits in resources and an increased focus on accountability, timely evidence-based decision making has become increasingly important in the policy area of family self-sufficiency and stability. Through collaboration with a research team, state and local agencies can seize opportunities to introduce an RCT into the normal course of program operations without the high cost or disruption to normal operations sometimes associated with large-scale RCTs. This lower-cost, minimally disruptive approach—sometimes known as *opportunistic experiments*—uses random assignment not only as a fair way to allocate access to a program or service, but also to generate strong and timely evidence that can support policy and program decisions and improvements.

An opportunistic experiment (OE) is a type of RCT that studies the effects of an initiative, program change, or policy action that an agency or program plans or intends to implement. In contrast, traditional, large-scale RCTs often examine an intervention or policy action that is developed and implemented specifically for a research study. OEs often use existing administrative data to measure outcomes for existing program participants. This can make them less costly and burdensome than other RCTs, which typically must collect new data and recruit participants. OEs initiated by state and local agencies can facilitate the recruitment of study sites, the cooperation of program staff, and the collection of low-cost administrative records data to measure outcomes. For these same reasons, findings from OEs can often be observed in a shorter timeframe. OEs typically focus on near- and intermediate-term outcomes typically observable within a 12 month period. By embedding research in planned changes and using existing administrative data, agencies can seize an opportunity to generate rigorous evidence on the effectiveness of program changes. As a result, OEs can help to inform ongoing program improvements in a timely manner.

A related evaluation technique—rapid-cycle evaluation (RCE)—can also be used to assess the effectiveness of program operations and services. While traditional RCTs and OEs exclusively rely on experimental (random assignment) designs, RCE may

use experimental or quasi-experimental evaluation designs (Box I.1). When RCE uses an experimental design, it may also be considered an OE. RCE is "rapid" because, like OEs, it focuses on relatively short-term outcomes, and because it relies primarily on the use of existing data to measure outcomes. The "cycle" in RCE refers to the use of evaluation in a continuous improvement framework—a series of iterative experiments to identify the relative effectiveness of business or program processes and service delivery approaches—whereby an agency makes incremental modifications to improve program effectiveness. In this toolkit we focus on the use of OEs in the policy area of family self-sufficiency and stability. Where relevant, we draw parallels with and provide examples of RCE.

BOX I.1

Rigorous Research Methods: Key Terms

Randomized controlled trial (RCT): An RCT is a rigorous research experiment considered to be the 'gold standard' for evaluating whether an intervention is effective in causing a given outcome. In an RCT, study participants are assigned, at random, to either a treatment group that receives an intervention, or a control group that does not. Random assignment ensures that the two groups are similar at the beginning of a study, which allows differences in the two groups' outcomes over time to be attributed to the intervention.

Opportunistic experiment (OE): An OE is a style of RCT in which an experiment is embedded into a planned initiative, program change, or policy action. OEs rely upon existing administrative data to produce findings in a quick turnaround and low-cost manner.

Rapid-cycle evaluation (RCE): RCE is an evaluation technique that uses either an RCT or a quasi-experimental ("comparison group") design to assess whether incremental changes in program operations or services cause an improvement in outcomes. RCE is "rapid" because it focuses on relatively short-term outcomes and relies on existing administrative data. RCE often goes a step further than an OE by using evaluation "cycles;" that is, an iterative approach in which programmatic changes are tested, modified, and tested again, and thereby are less "opportunistic" in nature.

This toolkit is designed to help researchers, policymakers, and agency administrators recognize and conduct OEs. In this first part (Module 1), we describe the use of OEs, explore their key features, and highlight several case studies. In the second part (Module 2), we discuss key steps and issues to consider when recognizing an opportunity for an experiment and determining the feasibility of conducting it. In the third part (Module 3), we outline and explain the critical steps in conducting an OE. This toolkit was produced as part of the Advancing Welfare and Family Self-Sufficiency Research Project (Project AWESOME), funded by the Administration of Children and Families' (ACF) Office of Planning, Research and Evaluation (OPRE) within the U.S. Department of Health and Human Services.[1]

[1] This guide on opportunistic experiments in family self-sufficiency and stability research is adapted from a recently published report in the education area: "Recognizing and Conducting Opportunistic Experiments in Education: A Guide for Policymakers and Researchers" (Resch et al. 2014).

HOW AND IN WHAT SETTINGS ARE OPPORTUNISTIC EXPERIMENTS USED?

OEs are common in a variety of settings. Fundraisers sometimes test different styles of messaging in their communications with donors and supporters. As an example of both an OE and an RCE, a recent presidential campaign embedded experiments in its emails to potential donors and supporters by randomly assigning different email messages to different groups on its mailing list (Green 2012; Issenberg 2012). The campaign then tracked how much money each group donated, determined which messages were most effective, and then used them for future appeals. Similarly, businesses test website designs to learn which ones works best in order to increase sales and improve user experiences (Mayer et al. 2009; Linden 2006). In education research, OEs frequently assess how educational strategies and practices influence student and school outcomes. For example, the New York City Department of Education recently embedded an OE in a planned literacy pilot intervention (Resch et al. 2014; Anand 2013; New York City Department of Education 2013). Forty middle schools were randomly assigned to receive the new literacy program or to serve as a control group. Existing student achievement data will be used to assess key outcomes of interest. If the new program improves reading achievement, the district will have rigorous evidence to support the expansion of the program to other schools; if it does not, the district can modify the program and then re-test it, or it can call off the expansion.

Government human services agencies also conduct OEs to learn how best to use scarce resources to better serve citizens. The Virginia Department of Social Services (VDSS) was charged with increasing participation in the Earned Income Tax Credit (EITC) program as a result of a recent policy directive. However, the most cost-effective mechanism to achieve this goal was unknown. Should the state send out letters to potentially eligible taxpayers? Or, would a phone call to notify likely eligible individuals of the program work better? Rather than assume which approach would be more effective in increasing program participation, VDSS took advantage of this opportunity to test and compare several outreach strategies and gather rigorous evidence to build its outreach strategy (Beecroft 2012). By randomly varying the outreach method across a set of households, Virginia learned that a strategy combining phone calls and mailings to eligible households increased EITC participation and made cost-effective use of state resources.

Oftentimes administrators are faced with implementing new policies or procedures without knowing if the changes will be an improvement over past practice. As another example of an OE in a human services agency, the New York City Human Resources Administration (HRA) used an experiment to test new administrative procedures to increase the establishment of child support orders for children receiving cash assistance (Dinan 2013). HRA determined that the new procedures, which minimized the required court appearances for noncustodial parents, were not successful at improving the rate of child support orders being established. The new procedure also took longer than the original procedures. Because of its study, HRA was able to clearly and quickly establish that the procedures were unsuccessful without having to implement them program-wide. In both the Virginia and NYC examples, the agency was able to undertake its experiment quickly and inexpensively by using existing administrative records and focusing on short-term outcomes.

Despite the relative ease of OEs, agencies often miss such opportunities in the family self-sufficiency and stability policy area. In such cases, an agency may continue to use resources—or even increase spending—on an intervention that is not shown to be effective. As an example, consider a state that invests in a new, more intensive job search assistance intervention for Temporary Assistance for Needy Families (TANF) clients. Without a test of the new intervention's effectiveness, the agency will not know whether and to what extent it may affect clients' employment status. Even if clients' employment outcomes improve over time, without randomly assigning clients, for example, as a new intervention is gradually rolled out, the agency will miss an opportunity to generate rigorous evidence of its effectiveness. While it might be tempting to attribute an improvement in employment outcomes to the new intervention, it is not possible to do so without a valid "counterfactual condition" that represents what would have occurred in the absence of the intervention. Without generating data to guide their decision, the agency may continue to fund the intervention without knowing whether that money is well spent.

WHAT ARE THE DEFINING FEATURES OF OPPORTUNISTIC EXPERIMENTS?

OEs typically have one or more of the following features, which may reduce costs or minimize the disruption to service delivery, and generate rigorous findings more quickly compared with traditional, large-scale RCTs:

- **An opportunity to randomly assign participants to an intervention.** In OEs, participants are randomly assigned to an intervention in a way that minimizes the disruption associated with random assignment for the entity being studied, as well as its participants (such as a local TANF agency and its clients). When random assignment is seen as a fair way to allocate limited resources and is planned in a way that does not cause major disruption, it is easier to randomly assign local agencies, cohorts of clients, or individual clients to treatment and control conditions.

- **Experimentation on a program or policy change initiated by the entity being studied.** OEs typically take advantage of an intervention or policy change that is initiated by the entity being studied. State and local agencies may implement new services or use new procedures in the belief that these changes will improve client (or other) outcomes, but lack rigorous causal evidence to support this belief. When administrators and researchers build research into a planned program or policy change, they introduce little disruption for participants and can evaluate whether the change does improve outcomes. Because the program or policy will be implemented with or without the evaluation, researchers working with agency staff may have an easier time recruiting sites. For example, if a state agency is planning to pilot a new program in a subset of its county agencies, researchers can work with state agency staff to encourage counties to participate in the study and undergo random assignment.

- **Strong partnerships between researchers and agency policymakers and administrators.** When research plans align with a state or local agency's programmatic plans, everyone can benefit. The state or local agency can help focus the research questions toward actionable results, and close participation of state and local admin-

istrators—prompted by their being invested in learning the findings—may also facilitate easier access to administrative data and enhanced cooperation of staff. If the OE is conducted by an internal party, the evaluation might require fewer clearances than traditional, large-scale RCTs, speeding up the timeline and reducing the disruption introduced by research. In Box I.2, we discuss the importance of close partnerships between researchers and both agency administrators and policymakers in identifying opportunities for research and conducting OEs (and other studies) with minimal disruption or added cost.

- **Potentially easy access to data on key outcomes.** The agency may already collect data on relevant outcomes through administrative records. Such readily available data may make a study less disruptive for participating agencies, staff, and clients than is the case in a study for which researchers must collect additional data to measure key outcomes (for instance, through a survey). An OE typically does, but need not have this feature. For example, an OE might capitalize on an opportunity to randomly assign participants to receive an intensive service intervention but require additional data beyond those collected conventionally (such as survey data on personal and family outcomes of participants).

BOX I.2

Close partnerships between researchers and agency administrators and policymakers

Ongoing, mutually beneficial partnerships between researchers and agencies can be critical to conducting rigorous OEs with little added cost or disruption. Some agencies may have a researcher or team of researchers on staff, which provides a natural opportunity for a partnership. If not, agencies may find a research partner by recruiting one from a local university, contacting a research organization, or reaching out to researchers who are conducting a study or collecting data in their state. Even prior to the initiation of any study, agencies that establish an ongoing "thought-partner" relationship with a researcher or team of researchers can foster a culture that supports evaluation and facilitates the early identification of research opportunities. Such thought-partner relationships can be:

- Established through regular lunch meetings or forums;
- Held with staff from several offices within the agency, rather than just one (given that formal interoffice contact may not be frequent enough to identify research opportunities sufficiently early); or
- Focused on sharing regular status updates, in which agency administrators and policymakers discuss planned initiatives and seek feedback from a research perspective, and in which researchers share results from their latest work and help identify potential opportunities to generate rigorous evidence on the effectiveness of an agency's initiatives

When an opportunity for experimentation has been identified, continuing communication between researchers and agency administrators and policymakers can help ensure that a research opportunity aligns with the agency's research agenda and is feasible to conduct. From there, close relationships between researchers and agency officials can help ensure that an experiment minimizes disruption for local offices, staff, and clients by facilitating recruitment, random assignment, and data collection, analysis, and reporting.

WHAT PROGRAM SITUATIONS PROVIDE OPPORTUNITIES FOR EXPERIMENTATION?

Three common situations can facilitate random assignment and the use of OEs with minimal disruption to participants—a randomized pilot, an oversubscription lottery, and a staggered rollout of a program change. In each of the situations below, random assignment is a fair way to allocate limited slots for an intervention in which there are limited resources. In this section, we briefly describe each of these situations, and highlight several case studies from recent research that are largely characteristic of OEs.

Randomized pilot. To test a new intervention, strategy, or program change before deciding whether to implement it more broadly, an agency can randomly assign sites, or individuals within a single site, to the pilot (treatment) group or to the business-as-usual (control) group.

Oversubscription lottery. When interest in or demand for an intervention exceeds capacity (oversubscription), agencies may use a lottery to randomly assign interested parties to either a treatment group (offered a spot in the program) or a control group (not offered a spot). Random assignment becomes a fair way to allocate access to a program or service.

Staggered rollout. When an agency lacks the resources to roll out a program change for all sites or individuals simultaneously, or simply wishes to understand the effect of the change before implementing it broadly, a staggered rollout design may be used to study an intervention or a change made to it. In this case, the treatment group may receive the program change during the initial phase of an experiment, while the control group essentially is placed on a wait list to receive it during a second phase. The second group can then act as a control to the former group during the first phase of the evaluation, but not after. If early analyses reveal that the program change is not effective, then the rollout in control sites can be cancelled or modified.

As an example of a pilot strategy implemented and tested in a staggered manner, consider a hypothetical situation in which a state's Supplemental Nutrition Assistance Program (SNAP) aims to increase program participation among youth and young adults formerly in the foster care system. To test a new outreach strategy for increasing participation, the state may pilot test the new strategy by staggering its implementation. The state can randomly select a set of local offices to implement it first and serve as the treatment group. The local offices not selected for the strategy essentially become the control group. If the study's findings show a short-term positive effect (for example, on program applications in the treatment group), then the state could implement the new strategy more broadly. If there are not positive effects, then the state may wish to modify the strategy and re-test it.

IN CONCLUSION: MAKING THE CASE FOR OPPORTUNISTIC EXPERIMENTS

The first part of this toolkit on the use of OEs in family self-sufficiency and stability research has shown that these relatively low-cost and quick turnaround experiments can be embedded within an agency's ongoing operations in a manner that provides strong and timely evidence of effectiveness. Agencies can seize opportunities to con-

duct experiments in cases in which the intervention being studied will take place with or without experimentation, where clients or sites can be randomly assigned with minimal disruption to agencies and their staff, where strong partnerships exist between the research team and agency administrators and officials, and where there is fairly easy access to data on key outcomes. Researchers can play a critical role in helping agency administrators and policymakers identify and conduct OEs and make evidence-based decisions related to family self-sufficiency and stability.

In the next two parts of this toolkit we discuss the issues to consider when recognizing an opportunity for an experiment and determining the feasibility of conducting it (Module 2), and then describe the key steps in conducting such an experiment (Module 3).

CASE STUDY 1

A supported employment pilot intervention in Ramsey County, MN

To improve paths to employment for hard-to-employ TANF recipients and their families, the Ramsey County, Minnesota, Workforce Solutions Department piloted an integrated service model known as Families Achieving Success Today (FAST). FAST was devised as a co-located partnership of several agencies to provide mental health, vocational rehabilitation, health care, and employment services to TANF recipients with disabilities. It was built on the successes of the Individual Placement and Support (IPS) supported-employment model for individuals with severe mental illness (Bond et al. 2008). FAST follows the core principles of the IPS model: finding competitive jobs in the community that fit participants' needs, strengths, and interests; fully integrating health services with employment services; and using rapid engagement into job search approach to help participants find jobs. Ramsey County partnered with a research team to conduct a small-scale random assignment evaluation of the FAST pilot to assess whether it improved client access to services, increased program and services uptake, and increased employment and economic stability outcomes (Farrell et al. 2013).[1]

From April to December 2011, 389 eligible TANF clients were randomly assigned to a treatment or control group. Treatment group members were offered a spot in the FAST program, while control group members continued to receive the usual case management and employment services accessible through TANF. The research team leveraged a variety of existing administrative data sources to measure impacts and other key research questions, including an automated case management system, case review notes, the state's benefit eligibility system, and unemployment insurance wage records. Qualitative interviews with key program staff and site observations were also conducted to examine fidelity of program implementation. The study found, among other things, that FAST treatment group members earned significantly more than control group members in the year after random assignment.

The FAST pilot exhibits characteristics of an opportunistic experiment. Ramsey County tested a program policy change using a random assignment design while maintaining business-as-usual services for control group families. This allowed the county agency to measure and assess impacts of the intervention against status quo service delivery. Using administrative data and the help of a small research team, the FAST pilot was rapidly rolled out and evaluated in the span of 16 months. A limited-scope qualitative study provided useful formative feedback and, coupled with the impact results, provided enough feedback to enable Ramsey County to continue services and expand the FAST intervention.

[1] This experiment was part of the TANF/SSI Disability Transition Project that tested innovative pilot interventions targeting TANF recipients with disabilities. The project was funded by the Administration for Children and Families within the U.S. Department of Health and Human Services, under contract with MDRC and MEF Associates.

CASE STUDY 2

A TANF and child support co-location pilot in Washington, DC

In 2005 and 2006, the District of Columbia (DC) implemented a pilot service delivery change to improve collaboration between its Child Support Services Division (CSSD) and Income Maintenance Administration (or TANF) and, in so doing, increase the number of child support orders. This initiative was funded via a section 1115 demonstration waiver grant from the Administration for Children & Families, which required a local evaluation of the intervention. To improve CSSD performance by increasing the number of TANF clients who participated in child support interviews (the first step in establishing a child support order), three CSSD intake workers were co-located at one TANF office. DC partnered with a research firm, the Lewin Group, to advise on and conduct an implementation study and impact evaluation of the pilot test (Farrell et al. 2007). More than 900 eligible clients were randomly assigned to a treatment or control group during a one-year period, where treatment clients received child support services on site at the TANF office. This involved taking part in both their TANF and child support intake interviews during a single visit to the TANF office. In contrast, control group members followed the standard procedure for child support intake, which involved an in-person interview at the separately-located CSSD headquarters.

Key data for the evaluation were collected from two existing sources: a baseline information form completed by all TANF applicants and the automated child support enforcement data system. Supplemental interviews and observations were conducted by the research partner to inform an implementation study. Outcomes were examined twice—at 6 and 12 months—following a client's eligibility determination and random assignment. The analysis showed that the treatment clients were significantly more likely to complete several key steps in the child support order process as a result of the co-location of the CSSD and TANF staff. The pilot test also demonstrated significant increases in the number of paternity and child support orders established for treatment clients compared with control clients.

This study has several characteristics of an opportunistic experiment. It was embedded in a planned service delivery change and utilized existing data sources to test impacts. Notably, data analysis early in the study identified that many treatment clients were not completing the child support interview following their TANF intake interview. DC quickly reordered the process such that the child support interview took place first, while clients were waiting for their TANF intake interview. This change largely resolved the issue and serves as an example of how data analysis during the early part of an experiment can help facilitate ongoing program improvement.

CASE STUDY 3

A lottery-based evaluation of an intensive coaching intervention for recovering substance abusers in the child welfare system in Illinois

Successive reports by the U.S. Government Accountability Office (1994, 1998) identified parental substance abuse as a contributing problem in more than two-thirds of active foster care cases nationwide. Access to appropriate treatment for substance-abusing parents is low, and only about one-third of those who need substance-abuse treatment receive it (SAMHSA 1997). In response, several states have undertaken experimental demonstrations of integrated child welfare and substance abuse service models using Title IV-E waivers, which allow states to bypass financing regulations and test strategies for better serving families and children through integrated care (Ryan et al. 2006). The Illinois Department of Children and Family Services (IDCFS) and Office of Alcoholism and Substance Abuse implemented one such Alcohol and Other Drug Abuse (AODA) Waiver Demonstration Project in April 2000 to test a model of intensive case management using substance abuse recovery coaches. The demonstration focused on two outcomes—improving access to substance abuse services and family reunification—over a three-year period from April 2000 to March 2003.

IDCFS leveraged its existing program infrastructure to support the demonstration. Eligible families were screened for substance abuse at regular Juvenile Court Assessment Program (JCAP) custody hearings. Upon receiving a referral for substance abuse treatment, families were randomly assigned to a treatment or control group. The control group received services as usual, while the treatment group received all standard services plus intensive case management delivered via a recovery coach. A total of 738 families, representing 1,417 children, took part in the study. The evaluation used several administrative data sources from the involved agencies to measure outcomes. The demonstration resulted in a significantly higher uptake of substance abuse services and a significantly faster entry into services among treatment families relative to control families (Ryan et al. 2006). Moreover, a significantly higher fraction of treatment families (12 percent versus 7 percent of control families) achieved parent-child reunification.

This study exhibits several characteristics of an opportunistic experiment. To increase interagency cooperation to improve program outcomes, IDCFS implemented a targeted service delivery change. The experiment leveraged the existing program infrastructure to carry out the intervention and administrative data to measure key outcomes. To address capacity issues, a lottery-style approach was used to randomly assign families to treatment or control groups. The study limited its scope by assessing two key outcomes for the intervention as a whole. This demonstration also laid the groundwork for a follow-on evaluation of the model's impact on substance-exposed births among families receiving recovery coaching (Ryan et al., 2008).

TOOLKIT MODULE 2

About this research guide: This document is the second in a set of three modules com prising the full "Toolkit on Recognizing and Conducting Opportunistic Experiments in the Family Self-Sufficiency and Sta bility Policy Area." Module 1 defines and describes opportunistic experiments (OE) and highlights several case studies. Module 2 outlines key steps in recognizing potential OEs and assessing their feasi bility. Module 3 provides an overview of the key steps in conducting an OE.

II. Recognizing Potential Opportunistic Experiments and Assessing Their Feasibility

In the face of limited resources and increased program accountability, human services agencies continue to invest in research that supports evidence-based policy and program decision making. Through collaboration with a research team, state and local policymakers and agency administrators can introduce an "opportunistic experiment (OE)" into normal agency operations by embedding a research experiment into a planned initiative, program change, or policy action. This relatively quick, low-cost evaluation approach can generate strong and timely evidence to inform policy and program decisions.

In this section of the toolkit (Module 2), we describe three key steps and related issues to consider when determining both the potential for designing and the feasibility of conducting an OE. The first step is to look for opportunities when program or policy decisions may be made randomly, such as those in the case studies described in Module 1 of this toolkit. Many situations in which a decision or choice is made are potentially opportunities to use random assignment. Once an opportunity for random assignment is identified, the next step is to consider the research and program questions of the state or agency and whether the potential evaluation aligns with these questions. The final step is to assess the feasibility of the evaluation. Integral to each step is a close relationship between researchers and agency administrators and policymakers (as discussed in Box I.2 in Module 1).

STEP 1. LOOK FOR POTENTIAL RANDOM ASSIGNMENT OPPORTUNITIES

In the usual course of program operations, certain situations can serve as opportunities to allocate program services through a random assignment process. Programs, services, or processes can be allocated through a randomized pilot, an oversubscription lottery, or a staggered rollout of a program change (as described in Module 1). Agency administrators and policymakers should consider how the decisions and choices they make may create an opportunity to use random assignment to allocate services and thus provide an opportunity for an experimental study. Common situations in human services agencies may include:

- **Pilot test.** An agency may be considering a new program or service delivery process but want to test the innovation before adopting it more broadly. State agencies may ask local agencies, for example, to volunteer to become pilot study sites. A randomized pilot would provide the opportunity to rigorously test the innovation. The state agency could randomly assign the pilot and control sites from the population of local agencies that volunteer.

- **Limited resources.** A state agency may be planning to introduce a new program enhancement or process change to all local agencies, but not have the financial or professional resources to implement the change in all local sites at once. This situation lends itself to a randomized rollout. Randomly assigning participants to differ-

ent phases of the rollout may not be overly disruptive, and it will allow the agency to evaluate the change.

- **Excess demand.** If a program or service is over-enrolled or is expected to have excess demand (perhaps due to limited resources), an oversubscription lottery can be used to assign slots. For example, in the education policy area, evaluations of charter schools and school vouchers have used a lottery mechanism to randomly assign students to overenrolled schools or oversubscribed voucher opportunities. In the area of family support and self-sufficiency, cases of excess demand may occur when new or a more intensive set of services are offered. For example, an agency may be offering new housing vouchers for low-income families or coaching and relationship education opportunities for child welfare and TANF clients.

- **Assignment of services.** In the normal course of program operations, administrators may at times have the opportunity to assign individual clients or local sites to contracted service providers or specific types of services. In some cases, these assignments can be made randomly, thereby creating an opportunity for an experiment. For example, a TANF agency could randomly assign clients or local sites to one of two employment service providers. In this case, an OE could be conducted to compare the approaches and mix of services delivered by the two different providers.

- **Centralized communication.** An agency may be trying to influence participant behavior through the centralized distribution of information—for example, by sending eligible families information on taking advantage of the EITC. If the agency wants to test the effectiveness of different strategies, an opportunistic experiment may be easier to implement if the communication effort is centralized. This can be viewed as a randomized pilot.

STEP 2. CONSIDER HOW OPPORTUNITIES ALIGN WITH PROGRAM AND RESEARCH NEEDS

While there are often potential opportunities for random assignment, before conducting an opportunistic experiment, it is important to consider whether it can answer an important research question. When making this decision, consider whether the research question relates to an established research or program agenda or supports an underlying need or priority, such as increasing client participation in an activity or improving employment outcomes. Moreover, it is important to bear in mind that different policy questions will be answered by different study designs. For example, randomly assigning program participants to different treatment providers may provide information on which provider is more (or less) effective, but it will not provide an answer to questions about the overall effectiveness of the program.

While addressing a previously identified priority may not be a prerequisite for the experiment, it is nevertheless important to consider the costs and potential disruption that may be incurred because of the experiment. Even relatively low-burden opportunistic experiments entail costs for agencies and their research partners, and not all opportunities to use random assignment for an experiment may be an efficient use of

limited resources. The agency should thoughtfully consider how experimentation can best be used to inform and improve program practices and service delivery.

STEP 3. ASSESS THE FEASIBILITY OF THE RESEARCH PROJECT

As in all research, the next step is to consider whether the research study is feasible. These six questions can be used for determining the feasibility of a study:

- What is the theory of change, and what comparison will be evaluated?
- What are the specifics of random assignment?
- What are the critical outcomes and data sources?
- Is adequate statistical power possible?
- Is research clearance needed?
- What is the study timeline?

In the remainder of this section, we examine each question in turn. Table II.1 highlights the key issues and questions, described below, in assessing the feasibility of a potential research project.

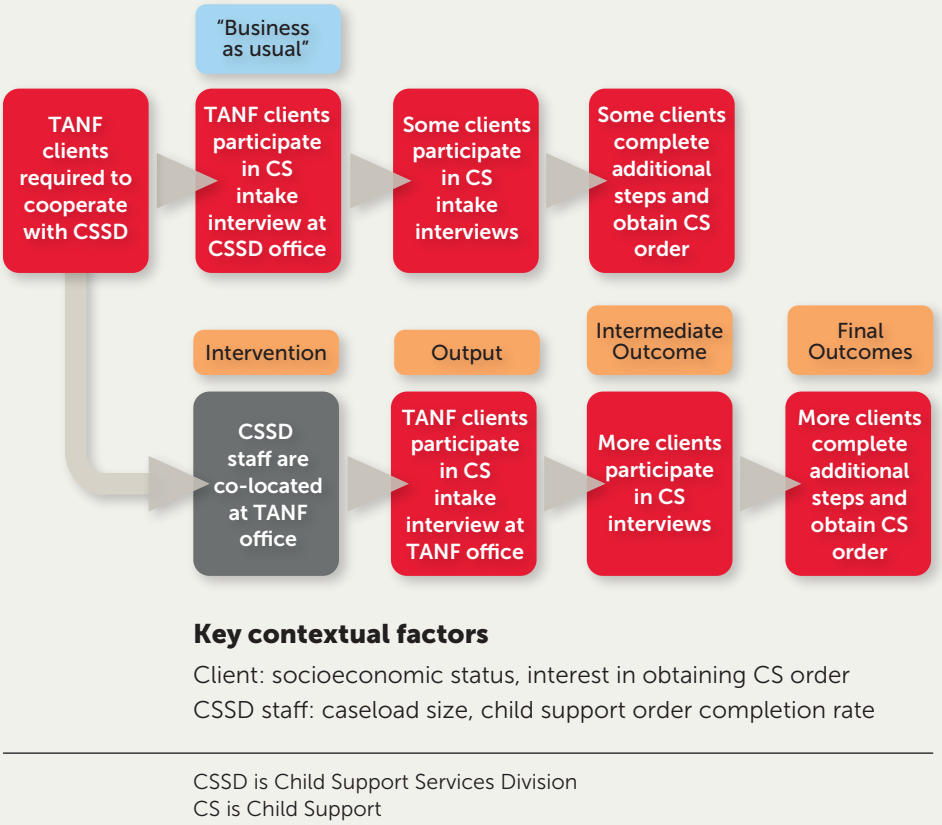
1. What is the theory of change, and what comparison will be evaluated? Before designing a research study, program administrators and researchers need to understand how an intervention works, consider who it will affect, and identify relevant outcomes. In other words, all evaluations of program interventions or process changes should begin with a theory of change.

A theory of change describes the causal chain of events that leads from the implementation of the intervention to the desired outcome. In some cases, the theory of change is simple—targeted GED tutoring and support is expected to directly influence the attainment of a high school credential—but in most cases it is more complex and thus helpful to create a diagram that captures the theory of change. This diagram includes the intervention, intermediate outcomes, and long-term outcomes, and it illustrates contextual factors that can affect the causal chain. Having a diagram provides structure for thinking about how the intervention affects long-term outcomes—or what actions or changes in behavior are expected to occur as a result of the intervention. If a theory of change does not exist, it is important to develop one, although the theory need not be a diagram. Researchers may need to lead development of the theory of change and then solicit feedback from the agency.

In Figure 1, we present a simple theory of change for the TANF and Child Support Co-Location Pilot Project in Washington, DC, described in Case Study 2 in the first part of the toolkit (Module 1). The top row of boxes shows the typical path of TANF clients in the absence of the intervention as they receive services through the Child Support Services Division (CSSD). TANF clients are scheduled for a child support intake interview and can receive related child support services through the CSSD office. Following the arrows, some TANF clients will participate in a child support

interview and, over time, some will establish paternity or obtain a child support order. The black box in the second row represents the intervention, that is, the co-location of CSSD staff in a local TANF office. The remaining boxes in this row show the anticipated results of the intervention— that more clients will actually participate in child support interviews and receive related child support guidance from the CSSD staff, thereby increasing the chances that TANF clients will complete additional steps in the process, establish paternity, and ultimately obtain a child support order.

Theory of change for the TANF and Child Support Co-Location Project in Washington, DC



CSSD is Child Support Services Division
CS is Child Support

Figure 1

A theory of change should explicitly differentiate between the proposed intervention (to which the treatment group is assigned) and current practice (to which the control group is assigned). In a research study comparing a new program with current practice, the study team refers to current practice as the "counterfactual condition"—that is, what would have happened to members of the treatment group in the absence of the intervention. In Figure 1, the top row represents the counterfactual condition. The second row shows the anticipated causal path when the intervention is implemented. A comparison of the two rows illustrates the key research question. In this case, the study examines the impact of the intervention (shown in the black box) on the outcomes in the boxes on the right of each row.

2. What are the specifics of random assignment? The second feasibility consideration is to determine the specifics of random assignment. Random assignment can happen in different ways, so researchers should consider the following key questions:

- **Who or what is randomly assigned?** The theory of change can help determine whether the appropriate level for random assignment is local sites, staff, and/or clients. If the intervention is an agency-wide program, and if it would not be possible to find a control group of clients within a local office that is unaffected in some way by the intervention, then random assignment should apply to local offices, rather than individual clients.[2]

  Consider how the level of random assignment is related to spillover, which refers to control group members possibly being affected by the intervention. For example, spillover may be a concern when rolling out agency-wide strategies if staff from different local offices interact regularly through professional events or intra-office work teams. In this way, the behavior of staff in the control group may be influenced by exposure to ideas or information from staff in the treatment group. Also consider how the level of random assignment is related to contamination, in which control group members actually receive all or part of the treatment. For example, in a lottery for an intensive intervention, the research team should closely monitor compliance with random assignment to ensure that clients assigned to the control group do not have access in any way to the treatment intervention.

- **Who is eligible to be randomly assigned? Are they representative?** In some opportunistic experiments, local sites or individuals may volunteer to be randomly assigned to treatment or control groups. They may be TANF clients wishing to enroll in a supported employment intervention, staff interested in a new training program, or local agencies willing to pilot a new program strategy. While randomly assigning volunteers does not affect the internal validity of a study, it can affect the generalizability of study findings. In some studies, such as a study of a supported employment intervention, the impact on volunteers may be part of the key research question. In other words, policymakers may want to understand the impact of the intensive intervention on clients who are motivated to participate in the intervention.

  In other cases, such as when an agency is piloting a new Supplemental Nutrition Assistance Program (SNAP) application process, policymakers may be more interested in the impact of the process on the average SNAP applicant, not its impact on applicants who choose to try the new process. The researcher must be aware of the implications of who is randomly assigned and interpret and present the results of the study accordingly. Likewise, it may help for researchers to provide agency administrators with examples of what a proposed study's results would—and would not—enable them to say about a program's effectiveness.

- **When and where in the service delivery process does random assignment occur?** If random assignment occurs at one point in time and in one location, such as by a central administrator after all applications to an intervention are submitted, it can be fairly straightforward to implement. If the intake for an oversubscribed inter-

[2] The level of random assignment also has important implications for the data analysis. For details, see Murnane and Willett (2011).

vention occurs on a rolling basis in multiple locations, such as at each local agency, random assignment may be more difficult to implement and monitor because more staff will need to be trained on the procedures.

- **How big will the treatment and control groups be?** Random assignment studies do not require that 50 percent be assigned to the treatment group and 50 percent to the control group. In lottery studies, the share of the participant pool assigned to the treatment group can vary widely on the basis of the amount of oversubscription. While the treatment–control ratio will have implications for statistical power (see the discussion of statistical power later in this section), other factors should also be considered. The number of treatment sites or individuals may depend on budget or capacity. For example, if an agency has limited funds to implement the intervention, it may choose to randomly assign only 30 percent of sites or individuals to the treatment group.
- **Will random assignment be implemented with integrity?** To ensure the integrity of experimental designs, compliance with random assignments must be monitored (and verified). In some cases, the researcher has control over compliance with random assignment, as when treatment sites receive tailored information that would not be useful to control sites, thus reducing the likelihood of spillover. In other cases, the researcher must monitor compliance with random assignment, as when monitoring a lottery process to ensure that lottery results are recorded accurately and that offers of intervention services are made only to those assigned to the treatment group.

3. What are the critical outcomes and data sources? The third way to assess the feasibility of a research project is to consider ways to reliably measure the key outcomes of interest and assess the difficulty of obtaining the required data. A key prerequisite to an efficient opportunistic experiment is the availability of relevant and usable administrative data with outcomes of interest that are observable in the data (Cody and Asher 2014). Using outcome measures from administrative records data may be the most cost-effective way to collect outcome measures for opportunistic experiments. In Box II.1, we highlight data often available through state agencies related to an individual family's self-sufficiency and stability. Although opportunistic experiments are often thought of as cheap and quick and thus focused on short-term outcomes, the need for long-term outcomes should not preclude a researcher from starting an opportunistic study. In Box II.2, we further explore long-term outcomes.

When assessing the potential outcomes and data sources, consider the following questions:

- **Are the outcomes cost-effective to collect?** It is important to be realistic about the resources necessary to evaluate impacts on outcomes of interest. With limited resources, evaluations may focus on a specific subset of outcomes that can be collected most easily.
- **Are the outcomes measured similarly for the treatment and control groups?** If not, the evaluation will not be able to determine whether differences in outcomes are due to differences in effectiveness or differences in how outcomes are measured.

- **Are the data complete and accurate?** What are the patterns of missing data; that is, are certain individuals more or less likely to have missing data? If outcome data are frequently missing or missing disproportionately for one group, the outcome may not be appropriate for evaluation or, at the least, results should be used with caution. In some cases, agencies may wish to supplement administrative data on employment, earnings, and other variables with data collected from a short survey, on-line or otherwise. A survey can be used, for example, to collect more detailed information on the characteristics of clients' jobs or to capture whether clients are working in informal jobs, out-of-state jobs, or self-employed jobs, which are not likely to be included in a state's administrative data.

BOX II.1

Key outcomes from state administrative data and other sources

Administrative data are an ideal source for outcome data in an opportunistic experiment. Human services programs like TANF, SNAP, and child support and child welfare programs track regular administrative data on client characteristics, program participation, and outcome data. In addition, unemployment insurance (UI) data from state departments of labor offer employment and earnings records data. In addition, many program administrators use readily available types of program-level data such as TANF work participation rates, SNAP error rates, and metrics for staff performance monitoring. It makes sense that when an opportunistic experiment tests for the effect of a program change (such as a more self-directed SNAP application process) on a measureable outcome (such as SNAP participation) to leverage existing administrative and program data capacities.

While the specific data elements vary by state, states' administrative and program data sets commonly include the key types of measures, below. Data sharing agreements between state agencies can facilitate a given agency's access to administrative data.

- Client identifiers (important for linking data sets)
- Client demographics and eligibility data
- Program enrollment, attendance, and participation
- Benefits received (in dollar values) for a given month or quarter
- Employment status in a given month or quarter
- Earnings (in dollar values) in a given month or quarter
- Child support processes and orders
- Child welfare services, processes, placements, and parent-child reunifications
- Child care, transportation, and other supportive service benefits received and amount in dollars
- Staff data

BOX II.2

Exploring long-term outcomes

Agencies may be less eager to conduct an opportunistic experiment if the key outcome of interest is not available for several years. However, the need for long-term follow-up should not necessarily preclude agencies and their research partners from starting such a study. Intermediate outcomes may provide early evidence of an intervention's success, and implementing random assignment preserves the validity of impacts on intermediate outcomes and options for future follow-up using administrative data or surveys.

In the case of the TANF and Child Support Co-location Project (Case Study 2 and Figure 1), for example, the ultimate goal of the co-location intervention was the longer-term outcome of increased child support orders for TANF clients. However, the study focused first on the key short-term outcome of client participation in the initial child support intake interview. The intervention's theory of change assumed that some clients did not participate in the intake interview because the location of the interview was not convenient. Documenting the short-term impact on client participation in the intake interview was the first indication that the intervention might have had an impact on the ultimate long-term outcome of an increase in child support orders. Short-term outcomes were measured six months after random assignment, while longer-term outcomes, such as the establishment of paternity and a child support order, were measured 12 months after random assignment.

4. Is adequate statistical power possible? The fourth feasibility consideration is whether the experiment will be able to detect an effect of a given size. Conducting a power analysis can help the research team determine how large a sample the opportunistic experiment will need to detect a policy-relevant effect with reasonable precision. Although researchers may not be able to control the sample size in an opportunistic experiment, they should consider statistical power analysis as a tool for determining whether the potential experiment can generate findings with adequate precision.[3] In some circumstances, such as in a state with a small number of local offices, it may not be feasible to evaluate a program implemented at the local office level.

Several factors can affect an experiment's statistical power:[4]

- **Size of the expected impact.** If an intervention is expected to have a large impact on outcomes, the impact will be easier to detect, which means that a smaller sample is required to detect that effect. Researchers should assess the size of the expected impact by considering the theory of change and examining prior research. Generally, an intensive intervention is more likely to have a large impact than a "light touch" intervention.

- **Sample size.** Statistical power generally improves with a larger sample size. Researchers should thus ensure that the estimated size of the research sample is sufficient to detect the expected impact. Box II.3 discusses how to proceed in cases when the sample size may be difficult to estimate. In studies using a rolling random

[3] For more details on statistical power analysis, see Murnane and Willett (2011, chapter 6).

[4] For more details on calculating statistical power for random assignment studies, see Schochet (2008).

assignment process, researchers will need to consider how long it will take to accrue the necessary sample.

- **Level of random assignment.** As mentioned above, statistical power improves with a larger sample size. However, if random assignment is conducted at the local agency level, the primary determinant of statistical power is the number of local agencies (sites), not the number of clients served by the local sites. In this case researchers should pay special attention to the number of local sites that can be included in the study.

- **Availability of baseline data.** Baseline data on characteristics related to the outcome can improve statistical power by accounting for variation in outcomes that can be attributed to differences in these pre–random assignment characteristics. Prior measures of the outcome—for example, food stamp receipt during the previous year—are especially valuable because they tend to explain a large share of the variation in the outcome measure (Bloom et al. 2007; Schochet 2008).

BOX II.3

Box II.3. Estimating the size of a research sample

If the random assignment mechanism is a lottery, it may be difficult to estimate the size of the research sample. The primary factor in determining sample size is the number of applicants or potential clients in the lottery. Examining historical applicant trends may provide some guidance, but researchers should also consider other factors that may affect demand for a program, including enrollment in similar or related programs, as well as social, demographic, and economic indicators in the community that may be suggestive of the potential demand for a program and/or the supply of eligible applicants. Programs may make modifications to a lottery process, such as exemptions for certain clients (for example, siblings or relatives of existing clients). These types of modifications should be considered when estimating the size of the research sample (Tuttle et al. 2012). See Box III.2 for more information on exemptions (exclusions) from random assignment.

5. Is research clearance needed? The fifth feasibility consideration is whether any research clearance is required. Research studies may be subject to multiple layers of clearance or review, depending on the details of the study and the policies of the participating states and agencies. Many states or agencies require researchers to complete a research application before undertaking a study. States and agencies may also require that researchers provide evidence of approval by an Institutional Review Board (IRB) or complete a data use agreement. The requirements can vary widely across states and agencies, but a close ongoing partnership between agency administrators, policymakers, and researchers can ensure an understanding of what is required and may lead to expedited review. Situations in which multiple layers of clearance and review are required may not be good candidates for opportunistic RCTs, because these reviews can be costly and time consuming. However, in many cases, an IRB clearance can be expedited for opportunistic experiments.

Research that involves human subjects may require approval from an IRB. An IRB review typically requires the researchers to present the research questions and plans for protecting the participants. There are state and local variations in requirements, but all research studies with human subjects are not required to go through the full IRB process. Studies viewed as posing minimal risk to human subjects may be exempted or eligible for expedited review.

A data-use agreement defines the terms under which researchers may access certain data that have been collected by another entity. Data-use agreements are usually required before external researchers can gain access to administrative records data. Such agreements help to ensure data confidentiality and data security, and prohibit the unlawful disclosure of personally identifiable information. The handling of data agreements varies by state. For example, some states provide researchers with confidentiality and data security restrictions, some states provide only identified data, and others require researchers to work onsite (or any combination of these scenarios).

Data-use agreements can also describe plans for the dissemination of study results and may specify the types of publications that would result from the study, such as public reports, issue briefs, or articles submitted to a peer-reviewed journal. Data-use agreements can specify, if requested, that the agency providing the data be kept anonymous. Agreements often grant the data-providing agency the right to review all publications for compliance with the agreement's term.

6. What is the study timeline? The final feasibility consideration is a matter of timing. As in many other research endeavors, the feasibility of an opportunistic experiment depends critically on the study timeline and whether researchers can implement random assignment at a time that works with the agency's decision-making calendar and other program priorities. Participation by the agency in another research study at the same time may also complicate its involvement and introduce additional burdens that make the new study prohibitive.

A second timing consideration has to do with results. The nature of the research questions asked in opportunistic experiments means that practitioners and policymakers want to know the results of the study quickly so that they can make decisions based on those results. Researchers should consider these interests while also being realistic about the study timeline. Factors that may affect the timeline include the need to secure outside funding, time to recruit study participants, and time to negotiate access to data. When results are needed quickly, the opportunistic experiment may still be a feasible option if researchers focus on only short-term outcomes.

Key issues and questions in assessing the feasibility of an opportunistic experiment

1. What is the theory of change?	• Can it explain the causal chain of events that leads from the implementation of the intervention to the desired outcome? • Is it explicit about how the proposed intervention is different from current practice?
2. What are the specifics of random assignment?	• Who or what is randomly assigned? • When and where does random assignment occur? • How big will the treatment and control groups be? • Will random assignment be implemented with integrity?
3. What are the critical outcomes and data sources?	• Are the outcomes cost-effective to collect? • Are the outcomes measured similarly for the treatment and control groups? • Are the data complete and accurate?
4. Does the experiment have sufficient statistical power to answer the questions of interest?	• What is the size of the expected impact? • What is the sample size? • What is the level of random assignment? • Are baseline data available?
5. Is research clearance needed?	• Does conducting random assignment or accessing outcome data require approval from an Institutional Review Board or a data-use agreement? • How can a research partner help facilitate the completion of these steps?

Table II.1

III. Conducting Opportunistic Experiments: Key Steps

TOOLKIT MODULE 3

About this research guide: This document is the third in a set of three modules comprising the full Toolkit on Recognizing and Conducting Opportunistic Experiments in the Family Self-Sufficiency and Stability Policy Area." Module 1 defines and describes opportunistic experiments (OE) and highlights several case studies. Module 2 outlines the key steps in recognizing potential OEs and assessing their feasibility. Module 3 provides an overview of the key steps in conducting an OE.

By embedding a randomized controlled trial (RCT) into a planned initiative, program change, or policy action, agency administrators, policymakers, and researchers can work together to conduct a quick turnaround, low-cost evaluation. Such an "opportunistic experiment (OE)" can generate strong and timely evidence to inform policy and program decisions. In this third section of the toolkit (Module 3), we provide an overview of five key steps that researchers, alongside the administrators and policymakers with whom they collaborate, can follow to conduct an OE. All experiments, whether considered opportunistic or not, include these steps: (1) identify sites; (2) encourage participation in the study; (3) conduct random assignment and monitor compliance; (4) collect data; and (5) analyze data and report findings. We focus on aspects of these steps that may be somewhat different in OEs than in traditional RCTs. We draw on examples, both real and hypothetical, to illustrate the key steps and related issues. Table III.1 highlights the various steps and issues, described below, in conducting an OE.

STEP 1. IDENTIFY SITES

In many opportunistic studies, the plans for the intervention determine the set of possible sites. This can be an advantage, because the researcher can focus immediately on the relevant sample. For example, in a hypothetical pilot of a new SNAP application process in a given state, the set of potential sites may be county offices in the state that are not already using the new application process. An opportunistic experiment with a well-defined set of possible sites may be more cost-effective because the researchers will require fewer resources for recruiting sites.

STEP 2. ENCOURAGE PARTICIPATION

In some cases, experiments do not require recruitment of sites or participants, as in instances of centralized communication (where participants do not need to agree to participate in the intervention) or oversubscription (where participants may have already agreed to participate in the intervention but there aren't enough slots). Where recruitment is necessary, opportunistic experiments still offer advantages. First, if sites are already planning to implement an intervention or are eager to implement it, researchers do not have to sell the study to sites (explaining how the benefits of the intervention outweigh the disruption that comes with it). Second, if the research relies on administrative data, the experiment does not require data collection that disrupts program practices, which may make it easier to recruit participants. Where random assignment is added for research purposes, the research team will need to work with the state, agency, or local sites that implement the intervention to encourage participants to undergo random assignment (see Box III.1).

STEP 3. CONDUCT RANDOM ASSIGNMENT AND MONITOR COMPLIANCE

Depending on what or who is being randomly assigned, the research team may work with administrators from the state or local agency to ensure that assignment to treatment and control groups is truly random. Three key issues to consider when conducting and monitoring compliance with random assignment include:

BOX III.1

Encouraging participation in research that uses random assignment

Program administrators may hesitate to randomly assign local sites or prospective participants to treatment conditions for many reasons, including ethical or logistical concerns (Cook 2001). The following strategies can address these concerns and encourage administrators and agency staff to participate:

- Emphasize that because the agency is planning to implement the intervention with or without a study, it is in their best interest to know whether the intervention works. An experiment can enable an agency to discontinue use of an ineffective intervention or expand use of an effective intervention.
- Name the study's potential benefits, such as enabling agencies to better serve clients, better allocate limited resources, and better use taxpayer dollars by continuing the use of effective interventions only.
- Discuss the ramifications of making decisions without rigorous evidence.
- Convey that in the case of oversubscription or limited resources (or both), randomization is a fair and transparent way to distribute resources and services.
- Assure participants that the study has been designed to minimize disruption for local sites, staff, and clients.
- When necessary, allow for a limited number of exemptions from random assignment.
- If needed, the agency can offer incentives, such as small gift certificates, to encourage participation.

- **Ensuring the system for assignment is truly random.** Where agencies have systems for assigning sites or individuals to a program or intervention, the research team should ensure that the systems are truly random. For example, some systems use methods that administrators describe as "effectively random" (such as assigning clients to a program by the alphabetical order of their names), which the research community does not typically recognize as random because the methods may result in systematic differences among groups.[5] In such cases, the research team should work with agency officials to agree on a truly random mechanism for assignment.

- **Accommodating agency needs and preferences for random assignment.** The research team may also need to work with agency officials to accommodate needs and preferences that influence random assignment. For example, in an opportunistic experiment in which clients are randomly assigned to case managers, researchers

[5] Effectively random methods may not meet evidence-based guidelines for random assignment and methodological quality if it is unclear that all units in the study were assigned by chance (U.S. Department of Education 2005).

can work with agency staff to accommodate preferences for caseload composition (such as a balanced level of client need across case managers).[6] Likewise, if an agency needs to roll out a program to only a few sites at a time due to resource constraints, researchers can conduct a randomized staggered rollout. Researchers and staff must identify any exceptions (exclusions) in advance of random assignment. In Box III.2, we discuss how researchers can accommodate exclusions from random assignment.

BOX III.2

Accommodating exclusions from random assignment

Just as in other randomized controlled trials, researchers may face pressures to include certain sites or individuals in the treatment group of an opportunistic experiment. One strategy to alleviate this pressure is to exclude some sites or individuals from randomization (and the study), but let them participate in the intervention. For example, staff may strongly prefer that a particular individual receive the intervention (such as a client nearing the end of a TANF time limit), so random assignment to the control group would be unacceptable. In such instances, these individuals could receive the treatment but not be included in the study.

To preserve both sample size and power, as well as the ability to generalize results to other settings, researchers should minimize such exclusions. When exclusions are necessary, researchers and agency staff should discuss and negotiate eligibility for exclusion from random assignment (and thus from the study) well before random assignment occurs. As in any study, research teams should work with agency administrators to keep accurate documentation of any exclusions.

- **Monitoring and confirming compliance with random assignment.** After conducting random assignment, researchers should work with staff to monitor and confirm compliance with random assignment so that no changes are made after assignments. The monitoring task for opportunistic experiments is similar to that for other RCTs: when relevant administrative records are available, researchers should verify that the actual assignments of local sites or individual clients to conditions match the original random assignment. Even in an oversubscription lottery for which the agency is using random assignment to regulate admission to oversubscribed programs, the researcher must still monitor random assignment and compliance. In Box III.3, we outline how researchers can ensure validity of randomization generated by lottery.

BOX III.3

Box III.3. Monitoring randomization in a lottery study

In lotteries for oversubscribed programs, attention to recordkeeping and compliance with the randomized list may vary across local sites. To ensure validity of randomization generated by lottery, the researcher should ideally:

- Participate in the lottery by observing or even generating the randomized lists.
- Compare the randomized lists to actual program offers and enrollment records.

[6] For example, researchers may pairwise match the research sample on relevant site or client characteristics when planning and conducting random assignment to ensure balance on key characteristics. After researchers have constructed the pairwise matches, sites or clients are randomly assigned within pairs (often using a random number generator) to treatment conditions.

STEP 4. COLLECT DATA

In some opportunistic experiments, all relevant data may be readily available (such as through various administrative sources), but others will require additional data collection. To preserve the possibility of longer term follow-up, researchers could collect individual identifiers, such as name, birthdate, address, and Social Security number. With this information, researchers may find sample members in other administrative data sources or contact them for a follow-up survey if funding for such research becomes available. However, these data can be difficult for researchers to collect because of privacy protections, and the collection of these data typically must be included in a data-use agreement.

STEP 5. ANALYZE DATA AND REPORT FINDINGS

Analysis and reporting activities are similar across opportunistic experiments and other RCTs. Researchers should conduct the same careful analyses that they would in any study, accounting for such issues as low treatment take-up, control group reactions caused by anticipation of treatment, spillover of the intervention to the control group, a lack of baseline data, and attrition.[7]

Disseminating the findings from opportunistic experiments, as well as the context in which the experiment took place, is extremely important. When disseminating findings, researchers should describe what is being compared with the intervention. It is much more informative, for example, to report that a job search assistance intervention improved employment-related outcomes relative to the existing program than to just describe the pilot as effective. Researchers can also emphasize that dissemination not only contributes important information to the field but also helps the agency and other stakeholders build an evidence base relevant to their context and disseminate the results to administrators and staff who are not involved in the study.

IN CONCLUSION

At every level of the human services system, leaders need to know which programs, policies, and practices are most effective as they allocate scarce resources. The three modules that comprise this report offer a toolkit to guide researchers, agency administrators, and policymakers as they work together to identify, plan, and conduct RCTs related to family self-sufficiency and stability. Relatively low-cost and quick turnaround opportunistic experiments can be embedded within an agency's ongoing operations in a manner designed to provide strong and timely evidence of effectiveness. Such studies allow agencies to learn about an intervention's effectiveness and make evidence–based decisions to support families as they work toward self-sufficiency and stability.

[6] Murnane and Willet (2001) and the What Works Clearinghouse guidelines (U.S. Department of Education 2011) provide guidance on how to account for these issues in analysis.

Key steps and questions in conducting an opportunistic experiment

Step 1: Identify sites	• What is the set of possible study sites?
Step 2: Encourage participation	• Are agencies already planning to implement the intervention or are eager to implement it? • Are there any incentives or benefits to the agency that can be used in recruiting sites and/or participants? • If needed, will state and local staff implementing and overseeing the intervention be able to work with the research team to encourage participants (either site-level or individual participants) to undergo random assignment?
Step 3: Conduct and monitor random assignment	• Does the research team need to work with agency administrators or program staff to ensure that the system for assignment is truly random? • Does the research team need to work with agency administrators or program staff to accommodate nonnegotiable needs and preferences in the random assignment process (such as balanced levels of need across sites)? • Will agency administrators or program staff be able to help researchers accurately document any exclusions? • Will agency or program administrators be able to help researchers monitor and confirm compliance with random assignment to ensure that no changes are made after generating assignments?
Step 4: Collect data	• Are all relevant data readily available (such as through various administrative sources), or is there a need to collect additional primary data? • When possible, will administrators or staff be able to help researchers obtain baseline and outcome data? • To preserve the possibility of longer term follow-up, will administrators or staff be able to help researchers collect individual identifiers, such as name, birthdate, address, and Social Security number?
Step 5: Analyze data and report findings	• Will administrators or staff be able to help researchers conduct the same careful analyses that they would in any study—accounting for issues such as low treatment take-up, control group reactions caused by anticipation of treatment, spillover, a lack of baseline data, and attrition? • To generate knowledge for the family self-sufficiency and stability community, will administrators or staff be able to help researchers report and disseminate findings from opportunistic experiments?

Table III.1

REFERENCES

Anand, A. "City Announces the 20 Schools That Will Have Longer School Days." Chalkbeat.org, June 10, 2013. Available at [http://gothamschools.org/2013/06/10/city-announces-the-20-schools-that-will-have-longer-school-days/]. Accessed July 15, 2013.

Beecroft, E. "EITC Take-up by Recipients of Public Assistance in Virginia and Results of a Low-Cost Experiment to Increase EITC Claims." Richmond, VA: Virginia Department of Social Services, unpublished manuscript, 2012. http://eric.ed.gov/?id=ED510037.

Bond, G.R., R.E. Drake, and D.R. Becker. "An Update on Randomized-Control Trials for Evidence-Based Supported Employment." *Psychiatric Rehabilitation Journal*, vol. 31, no.4, 2008, pp. 275–295.

Cody, S., and A. Asher. "Proposal 14: Smarter, Better, Faster: 'The Potential for Predictive Analytics and Rapid-Cycle Evaluation to Improve Program Development and Outcomes," in M.S. Kearney and B.H. Harris, eds., *The Hamilton Project: Policies to Address Poverty in America*. Washington, DC: Brookings Institution, 2014.

Cook, T.D. "Sciencephobia: Why Education Researchers Reject Randomized Experiments." *Education Next*, vol. 1, no. 3, 2001, pp. 62–68. Available at: [http://eric.ed.gov/?id=EJ637042]. Accessed June 2014.

Dinan, K. "Local Agency Lessons on Implementing Random Assignment: An Example from NYC's Child Support Program." Presentation to the Association for Public Policy Analysis and Management, Washington, DC, 2013.

Farrell, M., C. Anzelone, D. Cullinan, and J. Wille. "Taking the First Step: Using Behavioral Economics to Help Incarcerated Parents Apply for Child Support Order Modifications." OPRE Report 2014-37. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2014.

Farrell, M., P. Baird, B. Barden, M. Fishman, and R. Pardoe. *The TANF/SSI Disability Transition Project: Innovative Strategies for Serving TANF Families with Disabilities*. OPRE Report 2013-51, Washington, DC, Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, December 2013.

Farrell, M., K. Gardiner, L.J. Turner, and M. Mueller. "Evaluation of the District of Columbia's Child Support Co-Location Demonstration." Washington, DC: The Lewin Group, 2007.

Green, J. "The Science Behind Those Obama Campaign E-mails." *Bloomberg Businessweek*, Nov. 29, 2012. Available at [http://www.businessweek.com/articles/2012-11-29/the-science-behind-those-obama-campaign-e-mails]. Accessed July 15, 2013.

Issenberg, S. "The Death of the Hunch." *Slate*, May 22, 2012. Available at [http://www.slate.com/articles/news_and_politics/victory_lab/2012/05/obama_campaign_ads_how_the_analyst_institute_is_helping_him_hone_his_message_.html]. Accessed July 15, 2013.

Linden, G. "Early Amazon: Shopping Cart Recommendations," April 25, 2006. Available at [http://glinden.blogspot.com/2006/04/early-amazon-shopping-cart.html]. Accessed July 15, 2013.

Mayer, M., K. Hom, and J. Wiley. "Now You See It, Now You Don't." Google Official Blog, December 2, 2009. Available at [http://googleblog.blogspot.com/2009/12/now-you-see-it-now-you-dont.html]. Accessed July 15, 2013.

Murnane, R.J., and J.B. Willett. *Methods Matter: Improving Causal Inference in Educational and Social Science Research*. New York: Oxford University Press, 2011.

New York City Department of Education. "Chancellor Walcott and Speaker Quinn Announce Schools Selected for Extended Learning Day Pilot and Expansion of Middle School Quality Initiative," June 10, 2013. Available at [http://schools.nyc.gov/Offices/mediarelations/NewsandSpeeches/2012-2013/061013_qualityinitiative.htm]. Accessed July 15, 2013.

Office of Management and Budget. "Next Steps in the Evidence and Innovation Agenda." Memorandum to the heads of departments and agencies, authored by Burwell, S., C. Muñoz, J. Holdren, and A. Krueger. ". Washington, DC: Office of Management and Budget, Executive Office of the President, 2013. Available at [http://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-17.pdf]. Accessed July 15, 2013.

Resch, A., J. Berk, and L. Akers. *Recognizing and Conducting Opportunistic Experiments in Education: A guide for Policymakers and Researchers* (REL 2014–037). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Analytic Technical Assistance and Development, 2014. Available at [http://ies.ed.gov/ncee/edlabs.]

Ryan, J.P., S. Choi, J.S. Hong, P. Hernandez, and C. Larrison. "Recovery Coaches and Substance-Exposed Births: An Experiment in Child Welfare." *Child Abuse and Neglect*, vol. 32, no. 11, 2008, pp. 1072–1079.

Ryan, J.P., J. Marsh, M. Testa, and R. Louderman. "Integrating Substance Abuse Treatment and Child Welfare Services: Findings from the Illinois Alcohol and Other Drug Abuse Waiver Demonstration." *Social Work Research*, vol. 30, no. 2, 2006, pp. 95–107.

Substance Abuse and Mental Health Services Administration (SAMHSA). "Substance Use Among Women in the United States." Rockville, MD, 1997.

Schochet, P.Z. "Statistical Power for Random Assignment Evaluations of Education Programs." *Journal of Educational and Behavioral Statistics*, vol. 33, no. 1, 2008, pp. 62–87. Available at [http://www.mathematica-mpr.com/publications/redirect_PubsDB.asp?strSite=http://jeb.sagepub.com/cgi/reprint/33/1/62. http://eric.ed.gov/?id=EJ788461]. Accessed July 15, 2013.

Tuttle, C.C., P. Gleason, and M. Clark. "Using Lotteries to Evaluate Schools of Choice: Evidence from a National Study of Charter Schools." *Economics of Education Review*, vol. 31, no. 2, 2012, pp. 237–253. Available at [http://www.sciencedirect.com/science/article/pii/S0272775711001051. http://eric.ed.gov/?id=EJ957650]. Accessed July 15, 2013.

U.S. Department of Education. "Scientifically Based Evaluation Methods." *Federal Register*, vol. 70, no. 15, 2005, pp. 3585–3589. Available at [http://www2.ed.gov/legislation/FedRegister/finrule/2005-1/012505a.html]. Accessed July 15, 2013.

U.S. Department of Education. *What Works Clearinghouse Procedures and Standards Handbook (Version 2.1)*. Washington, DC: U.S. Department of Education, Institute of Education Sciences, 2011. Available at [http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v2_1_standards_handbook.pdf]. Accessed July 15, 2013.

U.S. Government Accountability Office. "Foster Care: Agencies Face Challenges Securing Stable Homes for Children of Substance Abusers." Washington, DC: GAO/HEHS Report #98-182, 1998.

U.S. Government Accountability Office. "Foster Care: Parental Drug Abuse Has Alarming Impact on Young Children." Washington, DC: GAO/HEHS report #94-89, 1994.

Weigensberg, E., C. Schlecht, E. Wiegand, S. Farris, C. Hafford. "Family Self-Sufficiency Data Center: Needs Assessment Report." Chicago: Chapin Hall at the University of Chicago, 2014.

OPRE

MATHEMATICA
Policy Research