



# Mini-RCTs: A Tool for Program Managers to Help Identify What Works

*Harry Hatry and Batia Katz*

*November 2022*

**Both public agencies and private nonprofit organizations (NPOs) across the United States seek to improve citizens' quality of life in dozens of ways. Managers, of programs big and small, make many decisions over the course of a year on how best to use their scarce resources.<sup>1</sup> Often, these decisions are not backed up by sound evidence demonstrating what service-delivery procedures work or work best. This brief describes a tool—mini randomized controlled trials (Mini-RCTs)—to help close this evidence gap for managers of almost any public service, and at a surprisingly low cost.**

Randomized controlled trials (RCTs) have been called the gold standard for identifying the impacts of public service-delivery procedures (Hariton and Locascio 2018). The key feature of an RCT is the comparison of outcomes for randomly selected customers who receive a new or modified service procedure with the outcomes of customers who have not received the new procedure. This provides strong evidence on the impacts of the new procedure.

RCTs have, in recent years, been encouraged by the federal government for evaluations of federal programs and services. Philanthropic foundations also have begun to push for RCTs to obtain stronger information about whether their funding works.

However, undertaking RCTs can be quite expensive, require highly specialized help, and take a long time to complete. This makes RCTs impractical for many public service agencies, especially small government agencies and most NPOs. These problems have deterred agencies in local, state, or federal, governments and NPOs from using the procedure (Buck and McGee 2015). Most agencies that have obtained adequate funds for RCTs have contracted out most of the RCT work.

Though the challenges may appear daunting, managers of any service can, with little or no external assistance, learn to apply the principles of RCTs to many services. They can use a highly simplified version of RCTs, which we are calling “Mini-RCTs.”<sup>2</sup> Results of Mini-RCTs are primarily used for internal agency decisionmaking and should not be viewed as clear evidence for external use (though the findings from some trials may be of interest to other organizations).

Overall, this brief encourages public service agencies and their managers to consider the opportunities and usefulness of undertaking Mini-RCTs to provide them with *higher*-quality information for improving the effectiveness, efficiency, and equity of their services, while recognizing their limitations.

## Three Key Principles of Traditional RCTs Adapted for Mini-RCTs

Below we identify three key principles of RCTs that are incorporated in Mini-RCTs.

1. Outcomes are compared between the service recipients<sup>3</sup> who received a new or modified service-delivery procedure (usually called the “treatment group”) and the service recipients who did not receive the procedure (usually called the “control group”).
2. The selection of clients into each group is done randomly. This is to minimize the likelihood that differences in outcomes between the groups occurred from causes other than the new or modified procedures.

For example, an agency might seek to improve staff performance by using a new type of training program. If most of the staff given the new type of training are staff with long tenure and most of the control group are new staff, it would be inappropriate to attribute better staff performance in the treatment group to the training.

3. During the trial period, conditions are controlled so that none occur that could significantly affect the validity of the comparisons.

For example, in the example of staff training above, if somehow a substantial number of staff in the comparison group received a similar training elsewhere during the trial period, this would contaminate the findings. You would not be able to attribute performance differences primarily to the training that was offered internally.

These core principles can be adapted to make RCT-style testing appropriate for managers of most services. The three key principles of RCTs, listed above, enable public service agencies to use a stepped-down version of RCTs (“Mini-RCTs”) for many public service operating decisions. They also help managers make decisions on many service-delivery approaches and are useful for organizations of almost any size, with little outside assistance.

Although rare, some Mini-RCTs of program service changes have been conducted in local and state government agencies, even without the benefit of modern software and hardware technology:

- In 2014, the School District of Philadelphia partnered with an outside research organization to determine if sending a postcard to parents would improve student attendance. The research organization randomly selected some parents to get the postcard and others who would not get the postcard. The low-cost RCT provided important evidence that the postcard was an effective method for improving attendance (Rogers et al. 2017).
- In 2015, the New Mexico Department of Workforce Solutions substantially reduced its unemployment overpayments, in part because of results from an inexpensive RCT testing the use of individualized pop-up messages in online application forms triggered by the identification of risks associated with a claimant's previous answers (PEW 2016). Some claimants were randomly selected to get individualized messages while others got generic messages.
- In 1996, the Minnesota Department of Revenue tested procedures to increase tax collection from taxpayers identified as having unpaid liability.<sup>4</sup> The agency randomly selected 100 such taxpayers to send a letter giving them the opportunity to resolve their liability before being billed and selected another 100 such taxpayers to whom it did not send a letter. The agency then compared the percentage of compliance by each group. It found that 75 percent of those sent the letter responded appropriately. Of taxpayers who were not sent the letter, only 25 percent paid their owed taxes. The cost of this experiment was quite low, including primarily only mailing costs and adding a code for each of the 200 taxpayers identifying which of the two procedures was used, and tabulating the outcome data for each group.
- The Lab @ DC and the DC Department of Human Services (DHS) worked together in 2017 to test the efficacy of a reminder letter for Temporary Assistance for Needy Families (TANF) program customers to renew their eligibility for benefits.<sup>5</sup> The Lab randomly assigned TANF families to receive either the standard renewal notices or the notices in addition to a reminder letter. They then compared the percentage of families who completed the renewal process between the control group and the test group. They found that 14 percent more families who received the letter finished the renewal process on time. This indicates that the letter, a low-cost intervention, could result in more families keeping TANF benefits. As a result of the study, DHS now sends this reminder letter to all households receiving TANF and has begun using a similar letter for other programs such as the Supplemental Nutrition Assistance Program (SNAP).

This brief provides public service managers and their staff members with information and guidance for using stepped-down RCT processes to improve their decisionmaking about many agency operational service-delivery procedures. The following sections discuss these topics:

- benefits of using Mini-RCTs
- examples of categories of service-delivery procedures that can use Mini-RCTs

- the prerequisites needed to make Mini-RCTs feasible and appropriate
- the cost and staffing needed for administering the trials
- the limitations of Mini-RCTs

In addition, the appendix suggests the basic steps for starting and running a Mini-RCT.

## Benefits of Using Mini-RCTs

Mini-RCTs have substantial promise for helping program managers make service-delivery decisions.

Box 1 lists potential direct and indirect benefits.

---

### BOX 1

#### Benefits of Mini-RCTs

- Mini-RCTs enable new untested service-delivery procedures to be tested on their likely impacts before the procedures are implemented across the board, reducing uncertainties;
- test smaller-scale iterative improvements to current service-delivery procedures;
- encourage considerably more innovation and creative thinking about service-delivery approaches;
- help build organizations' capacity to become better learning organizations;
- increase interest in and use of RCT principles and reduce fear about RCTs; and
- impress and gain support of funders by providing stronger evidence of successful outcomes.

All of the above lead to more effective, efficient, and equitable services for customers.

---

## Examples of Service-Delivery Procedures That Can Use Mini-RCTs

Mini-RCTs can help address many categories of issues. We described four examples of real-life trials of this kind in the earlier section on key principles of Mini-RCTs. Further examples are listed in box 2.

We focus primarily on human services programs in this brief, but Mini-RCTs can also be effectively used to assess many other types of public services. For example, Mini-RCTs can also be used to test different procedures for maintaining roads, transit systems, public safety, and water and sewer systems. Note that in these cases important outcomes sought in the analysis might be the condition of physical units, rather than improvements in the condition of human beings, such as the condition of roads, buses, and water and sewer pipes.

---

## BOX 2

### Potential Service-Delivery Procedures That Can Be Tested by Mini-RCTs

- Mini RCTs can test alternative ways of delivering services such as counseling or other client treatment methods;
- emerging new technologies that seem likely to improve service effectiveness and/or efficiency;
- alternative lengths, intensity, timing, and number of client contacts provided by the service;
- approaches to raising program revenues;
- ways to recruit customers or volunteers;
- ways of effectively communicating about the program with clients or the public;
- improving the quality and cost of internal supportive services such as public-service agencies' own housekeeping, personnel, vehicle maintenance, and IT services; and

Mini-RCTs can also be used to shed light on equity issues by comparing the outcomes of a program by individual demographic subgroups (such as race/ethnicity, gender, or age group).

---

## Prerequisites

Mini-RCTs can be helpful in a variety of settings and for many types of programs. However, using Mini-RCTs has some prerequisites:

1. The operating environment and the new, or modified, service-delivery procedure that will be compared with the existing procedure is likely to remain reasonably stable throughout the trial. For example, no budgetary changes are planned that might substantially affect the program during the trial period.
2. It is possible to assign participants randomly into treatment and control groups and keep them in those groups for the duration of the trial.

Participants of many human service programs should be able to give their informed consent for whether they agree to participate in the trial.

Additionally, there should be limited possibility for spillover effects. Especially with new procedures, it is important that if participants are randomly assigned, staff should ensure they are using the correct procedures with each group. If staff are randomly assigned, it is also important that staff do not discuss the trial with one another.

3. The program has the capacity to collect and analyze the needed outcome data and to quickly apply the findings to changes in practice.
4. Important outcomes are expected to occur within a relatively short amount of time, such as within six months after service delivery.

5. No ethical problems are expected to arise from limiting the new procedures to only the test group. For example, RCTs are not likely to be appropriate if evidence already exists that the new procedure is effective in achieving important outcomes, which would make it unethical to deny the procedure to the control group while giving it to the test group.
6. The number of clients expected to enroll in the service during the test period is enough to provide clients for each of the experimental and control groups.

## Cost of Mini-RCTs

Unlike regular RCTs, Mini-RCTs are low cost, both in terms of dollars and staff time. As indicated in the steps described in the appendix, most programs should be able to train existing staff to organize and monitor the Mini-RCT, without having to hire additional help. There are costs associated with some Mini-RCT technical tasks, mostly for the initial Mini-RCTs trials. These include some staff training and setting up a computer system to process and analyze the comparison data.

Outside assistance will likely be needed, at least until experience has been gained with Mini-RCTs, such as to review each Mini-RCT's overall design, check questionnaires used to obtain outcome information, and help design the randomization procedure. Current hardware and software technologies make many of these tasks easier.

Here are some suggestions to keep costs low:

- If you already have a process for creating program performance reports, build on them for many of the technical tasks.
- Use email programs such as Qualtrics to obtain outcome information from customers, clients, or patients.
- Do not require unnecessarily high levels of statistical precision that would call for larger-than-needed sample sizes.
- Find internal staff interested and experienced in data and data processing to set up the data collection procedures and check the quality of the survey questionnaire used to collect data on program outcomes.
- Use outside “volunteers” (such as college or university faculty and students) for help, such as on randomization procedures and how to deal with procedure challenges that arise.

## Limitations of Mini-RCTs

Although Mini-RCTs are significantly less expensive than full-scale RCTs, they do have significant limitations:

- Mini-RCTs are not adequate for testing major complex program or policy options, especially those with major long-term outcomes or large amounts of funding.
- The findings will primarily be useful to the individual organization and program, not to other similar programs in other organizations.
- To the extent that the program has only a small number of customers, the accuracy and reliability of the findings will be reduced. This becomes more of an issue for comparing subgroup outcomes of different customer subgroups, such as different race, ethnicity, age, or gender subgroups, especially if more than two comparison groups are examined.

For example, a Mini-RCT might involve 150 customers and compare success in finding jobs using a new way of contacting employers. It would involve about 75 customers in each of two comparison groups. However, if the program wanted to compare success rates for three age groups, about 25 customers from each age group would be compared. The subgroup comparison findings provide a lower level of confidence than for the aggregate numbers.<sup>6</sup> This could have significant implications for studying and increasing equity.

Nonetheless, a good case can be made that the data obtained in these much simpler Mini-RCTs are likely to yield accurate enough data, given the considerably less-complex questions Mini-RCTs address. Even small agencies with a small number of small programs could likely implement a stripped down RCT and learn a lot about what works and does not work. For making many operating decisions, Mini-RCTs should be able to provide stronger evidence than otherwise is available to the program—and at low cost. To paraphrase from a common saying, it is better to be roughly right than precisely ignorant.

## Conclusion

We hope the above material will inspire public service organizations of all sizes and service types to test new or modified procedures to improve decisionmaking using Mini-RCTs. Public service organizations with little-to-no external assistance can likely learn to apply RCT principles to help improve the outcomes of their programs—and at low cost.

These stepped-down Mini-RCT procedures have important limitations in what they can provide and are not appropriate for examining major program and policy questions. These procedures do not seek to examine blockbuster innovations that require sophisticated, highly rigorous analysis. However, public service managers make many decisions each year about their operating procedures.

Mini-RCTs have the potential to yield many small-scale benefits that add up to major national improvements in the effectiveness, efficiency, and equity of public services—by providing more accurate information on what service-delivery practices work and do not work. A major additional benefit is that Mini-RCTs can stimulate innovation and creativity by giving decisionmakers at organizations the option of testing changes before fully implementing them. Ultimately, Mini-RCTs in public service program management could broadly benefit service delivery and help customers across all types of organizations.

## Appendix. Steps for Implementing a Mini-RCT

Each Mini-RCT will be unique to the service it evaluates and the staff conducting the trial. However, some basic steps apply to most, if not all, Mini-RCTs:

1. Select the new or modified service procedures to be compared. Thoroughly define the characteristics of the new procedures. (If feasible, use a pilot period to try out the new procedure ahead of the trial. This would catch some problems, making the new procedure more stable during the trial.)
2. Select the outcome indicators for comparing the effects of each comparison group.

This choice will be based both on the nature of the data you are already collecting and what new performance indicator data can be obtained at an affordable cost. Indicators might include data already being collected, such as data on changes in customer knowledge, behaviors, or conditions. For example, staff could measure the percentage of customers satisfactorily completing a training program or percentage of customers correctly completing applications for a service. Other more important indicators could be post-service outcomes, such as the percentage of customers that are employed six months after receiving job training, obtained by surveying participants.<sup>7</sup>

Mini-RCT performance indicators seldom measure the ultimate outcomes sought (which can be very expensive to obtain). However, programs should be able to identify “intermediate” outcomes that can be measured and indicate meaningful progress toward the end outcomes sought. For example, in the example discussed above, the percentage of customers completing a training program to help them find jobs is a low-level outcome indicator useful to program management. However, it does not provide data on the ultimate goal of finding long-term jobs for the unemployed. The program might even be able to obtain data on the percentage of those customers that were employed, say, three or six months after completing service.<sup>8</sup> Data on both training completion and short-term employment would help the program track progress and indicate where problems arise.

3. Choose a randomization method for deciding which incoming participants will be served using the new approach and which will be included in the comparison group. The randomization procedure might be as simple as using one of many free random number generators available online.

Train staff about random assignment. Seek staff engagement and buy-in for the trial.

4. Choose the length of the trial, considering the desirability of having as many participants as you can afford and the time when the findings from the completed trial will be useful. The larger the number of participants, the more accurate the findings. The method should select a representative sample of participants for each approach. For example, with random assignment and a large enough sample size, approximately the same percentage of difficult-to-help participants will be included in each group.<sup>9</sup> Randomization helps assure the comparisons will be valid, considerably increasing the strength of the evidence.



5. As each participant enters the program, assign them to one of the two groups using the randomization procedure the program has identified. Record which participants are assigned to which service approach.

Compare the demographics of the people in the treatment versus control group and potentially rerandomize the groups if they do not look similar, to better ensure the groups are representative of participants.

6. Select the total number of participants for inclusion in the trial. The larger the number, the more accurate the findings will be, but the cost of data collection will be higher.

These Mini-RCTs are not intended for use in complex, high-stakes programs that seek high levels of rigor and precision.

7. For some human services programs, participants should be told about the trial as they enter the program and asked to give their consent to participate in the trial, regardless of whether they are in the treatment or control group. For these programs, no one who has not given consent should participate in such Mini-RCTs.

8. Track the outcomes for each participant in each group at regular intervals.

9. Throughout the test period, adhere as closely as possible to the key characteristics of the service-delivery procedures being compared.

For example, if during the test period situations arise that substantially change service delivery or if, for any reason, substantial numbers of customers choose to obtain services elsewhere, this can weaken the accuracy of the findings and even invalidate the comparisons.

10. Calculate the values of each outcome indicator for each tested service-delivery approach and compare these values.

11. Examine the data, looking for other factors that might have affected the findings.

For example, if a public works agency is using a Mini-RCT to determine the material to use in maintaining roads, if unexpected weather conditions occur, this might explain the test outcomes.

12. Use the findings to adjust program practices.

The program might drop the approach that shows the poorer outcome. Or the information obtained might suggest the need for particular changes to the service delivery procedures. Alternatively, staff might decide that they have not yet obtained clear enough information from the outcome data, in which case they might continue the comparison for a longer time. Also, sometimes the approach that showed the poorer findings might be the better approach if it had more resources. As in any evaluation, it is important to look for qualitative information as well as quantitative data indicating reasons for the outcomes.

13. Communicate the findings with program stakeholders.

14. Use advances in data visualization techniques to display the findings and communicate and disseminate the information more effectively.

## Notes

- <sup>1</sup> This brief builds on previous Urban Institute work on building evidence into program management processes. To read the earlier briefs in the series, see Hatry et al. (2022) and Hatry (2020).
- <sup>2</sup> This brief seeks to build on recent work done in the field of behavioral economics (e.g., Richburg-Hayes et al. 2017). This brief also seeks to build on work on rapid cycle evaluations (e.g., Atukpawu-Tipton and Poes 2020). Both of these movements encourage low-cost and quick turnaround time evaluations, this brief shares these objectives. However, we also make the case that public sector programs can themselves adapt key elements of the much more sophisticated—and more costly—procedures of full RCTs to test new program service-delivery interventions before they are implemented across the board.
- <sup>3</sup> RCTs might also test the outcomes of others besides service recipients, such as regulated parties, staff members, and potential partners.
- <sup>4</sup> More information can be found on page 54 of Liner et al. (2001).
- <sup>5</sup> More information on this project, and others conducted by The Lab @ DC can be found on its website: [thelabprojects.dc.gov](https://thelabprojects.dc.gov).
- <sup>6</sup> This sample size limitation sometimes can be alleviated by combining subgroups.
- <sup>7</sup> Most of these postservice outcomes would be measured at 3 or 6 months after the end of service. After about 12 months it will likely be increasingly difficult and costly to obtain important outcome information, if not available from administrative records.
- <sup>8</sup> This type of information may be difficult to get postprogram but could be collected through a survey or data match that participants consent to ahead of time.
- <sup>9</sup> A large enough sample size can vary depending on the program or procedure being tested, but we suggest a sample of at least 100 participants, with about 50 participants in each of the treatment and control groups.

## References

- Atukpawu-Tipton, Grace, and Matthew Poes. 2020. "Rapid Cycle Evaluation at a Glance." OPRE Report #2020-152. Washington, DC; Arlington, VA: US Department of Health and Human Services (HHS), Administration for Children and Families (ACF), Office of Planning, Research, and Evaluation (OPRE); James Bell Associates.
- Buck, Stuart, and Josh McGee. 2015. *Why Government Needs More Randomized Controlled Trials: Refuting the Myths*. Houston: Laura and John Arnold Foundation.
- Hariton, Eduardo, and Joseph J. Locascio. 2018. "Randomised Controlled Trials—the Gold Standard for Effectiveness Research." *BJOG: An International Journal of Obstetrics and Gynecology* 125 (13): 1,716. <https://doi.org/10.1111/1471-0528.15199>.
- Hatry, Harry P. 2020. *Evaluation Guide for Public Service Program Managers*. Washington, DC: Urban Institute.
- Hatry, Harry P., Batia Katz, Katherine Barrett, and Richard Greene. 2022. *Dos and Don'ts: Tips for Strengthening Your Performance Management Systems*. Washington, DC: Urban Institute.
- Liner, Blaine, Harry P. Hatry, Elisa Vinson, Ryan Allen, Pat Dusenbury, Scott Bryant, and Ron Snell. 2001. "Making Results-Based State Government Work." Washington, DC: Urban Institute.
- PEW (PEW Charitable Trusts). 2016. "Behavioral Analytics Help Save Unemployment Insurance Funds: New Mexico Uses Data to Identify Misinformation, Save Money." Washington, DC: PEW.
- Richburg-Hayes, LaShawn, Caitlin Anzelone, and Nadine Dechausay with Patrick Landers. 2017. *Nudging Change in Human Services: Final Report of the Behavioral Interventions to Advance Self-Sufficiency (BIAS) Project*. OPRE Report 2017-23. Washington, DC: HHS, ACF, OPRE.
- Rogers, Todd, Teresa Duncan, Tonya Wolford, John Ternovski, Shruthi Subramanyam, and Adrienne Reitano. 2017. *A Randomized Experiment Using Absenteeism Information to "Nudge" Attendance*. Washington, DC: US Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

## About the Authors

**Harry Hatry** is a distinguished fellow at the Urban Institute, where he has been a leader in developing performance management/measurement and evaluation procedures for public agencies since the 1970s. He has worked with a wide range of local, state, and federal agencies—internationally and nationally—to develop outcome measurement procedures for such services as public safety, health, transportation, education, parks and recreation, social services, environmental protection, and economic development. He is author of the Urban Institute book *Performance Measurement: Getting Results*, second edition.

**Batia Katz** is a research analyst in the Center on Labor, Human Services, and Population, where she researches workforce development. Katz's previous research experience includes studying the science labor market, the impact of personality traits on employment outcomes, and gender and family in the workforce. Katz graduated with high honors from Haverford College, where she earned a BA in economics.

# Acknowledgments

This brief was funded by the Annie E. Casey Foundation and the William T. Grant Foundation through the Urban Institute's Federal Evaluation Forum. We are grateful to them and to all our funders, who make it possible for Urban to advance its mission.

The views expressed are those of the authors and should not be attributed to the Urban Institute, its trustees, or its funders. Funders do not determine research findings or the insights and recommendations of Urban experts. Further information on the Urban Institute's funding principles is available at [urban.org/fundingprinciples](https://urban.org/fundingprinciples).

Thank you to Shelley Metzenbaum, Katherine Barrett, Richard Green, Devlin Hanson, Stephanie Shipman, and Timothy Triplett for their thoughtful reviews of this report and helpful suggestions. We also thank Liza Hagerman for her editorial assistance.



500 L'Enfant Plaza SW  
Washington, DC 20024

[www.urban.org](https://www.urban.org)

## ABOUT THE URBAN INSTITUTE

The nonprofit Urban Institute is a leading research organization dedicated to developing evidence-based insights that improve people's lives and strengthen communities. For 50 years, Urban has been the trusted source for rigorous analysis of complex social and economic issues; strategic advice to policymakers, philanthropists, and practitioners; and new, promising ideas that expand opportunities for all. Our work inspires effective decisions that advance fairness and enhance the well-being of people and places.

Copyright © November 2022. Urban Institute. Permission is granted for reproduction of this file, with attribution to the Urban Institute.