



Strategies for RESEA Evaluation in Small States

Jacob Alex Klerman and Victoria Wheat (Abt Associates)

May 2023

The U.S. Department of Labor (DOL) strongly encourages all states to conduct rigorous impact evaluations of their Reemployment Services and Eligibility Assessment (RESEA) programs. Compared to states with large RESEA programs, small states with small RESEA programs face a unique challenge in conducting these impact evaluations.

This brief describes the challenge that small states face and identifies four strategies to help small states and their evaluators to address the challenge and maximize learning (see Box 1). Some strategies in this brief are designed to help states to increase their sample sizes while other strategies aim to reduce the sample size needed.

THE SMALL STATE CHALLENGE

Small states administering RESEA programs share a common challenge: achieving adequate sample sizes to detect impact on outcomes for claimants. To help states address this challenge, it is important to first discuss why such large sample sizes are necessary. In short, research has shown that RESEA program impacts tend to be small (see Box 2). Program impacts are small because RESEA is a low-intensity program. **To detect small impacts, large sample sizes are needed.**

Achieving large enough sample sizes to detect impacts in small states, even when those impacts are present, can be challenging for states with smaller populations and fewer RESEA-eligible claimants. There are strategies and approaches that small states can use to address this challenge and achieve an appropriate sample size—the adequate sample size to reliably detect impacts which includes both the treatment and control groups. Identifying the appropriate sample sizes to reliably detect impacts in a small state and laying the groundwork to achieve those appropriate sample sizes can be done early in impact evaluation planning, even before hiring an evaluator.

Box 1: Strategies Presented in This Brief

1. Increase your state's sample size.
2. Select a research design to reduce the appropriate sample size.
3. Choose research questions with smaller appropriate sample sizes.
4. Combine evidence from your state with evidence from other states.

Box 2: Past Research on RESEA Program Impacts

RESEA programs that aim to help Unemployment Insurance (UI) claimants move to employment are typically low-intensity, consisting of a limited number of one-on-one meetings or group sessions to provide job search assistance (Mills de la Rosa et al., 2021). Low-intensity programs, like RESEA, are expected to have small impacts on outcomes, such as UI benefit receipt and employment and earnings. For example, evaluations of the Reemployment and Eligibility Assessment (REA) program (predecessor to RESEA) showed impacts of reduced UI benefit duration by only 0.5 to 1.5 weeks and raised short-term (Q2) employment and earnings by only 1 to 3 percentage points (Klerman et al., 2019).

This document was prepared for the U.S. Department of Labor (DOL), Chief Evaluation Office by Abt Associates under Contract # 1605DC-18-A-0037. The views expressed are those of the authors and should not be attributed to DOL, nor does mention of trade names, commercial products, or organizations imply endorsement of same by the U.S. Government.

The starting point is to understand how large a sample size must be to reliably detect impacts when present. Earlier impact evaluation analyses suggest that to detect an impact of a RESEA program on *UI benefit duration* (weeks of UI received), a state will need a sample of roughly 10,000 UI claimants. These analyses also suggest that much larger sample sizes, likely 30,000 or more claimants, are appropriate to detect impacts on *Q2 employment* (employment status in the second quarter) after the start of a UI claim. For evaluations of RESEA *program components*, such as enhanced services or subsequent meetings, appropriate sample sizes are typically even larger (Klerman et al., 2019).

Even with these numbers in mind, how can a RESEA program in a small state attain samples large enough to detect impacts? With the right knowledge and support, even small states can conduct informative impact evaluations. The next sections provide four overarching strategies and actions any state, and particularly small states, can use to achieve appropriate sample sizes for RESEA evaluations.

1. INCREASE THE STATE'S SAMPLE SIZE

RESEA programs can take the following three actions to increase the sample size for evaluations.

Involve the whole state in the evaluation.

Often due to logistical reasons, a state might consider conducting an evaluation in only a select subset of those local areas or offices where the RESEA program operates. Particularly in small states, omitting some local areas or offices can exacerbate issues with inadequate sample sizes. To increase a state's sample size, one option is to include as many offices as possible from the RESEA evaluation. When exclusions are required, it is important for the state to include offices that serve large numbers of RESEA participants.

This approach also helps a state learn about the entire RESEA program impact. If parts of a state are systematically excluded, the evaluation might not reflect the program as implemented across the state. Because effectiveness could vary across local areas or offices, leaving some offices out of the study means the results will not reflect the average statewide impact of the program. Understanding of the RESEA program's overall impact is the best way to determine how to improve services.

Conduct the evaluation for as long as possible.

It can take a significant amount of time to enroll a sample adequate to detect impacts. Larger states might be able to build a large enough sample by enrolling new UI claimants for a year, but a smaller state is likely to need to enroll new claimants for several years. A RESEA evaluation can be designed to produce other valuable information (such as the implementation of program services) in the shorter term, but definitive information on final impacts will take as long as required to build an appropriate sample. Setting a realistic timeline is a crucial consideration in evaluation planning and in procuring a third-party evaluator.

Involve all RESEA-eligible UI claimants in the evaluation.

Until recently, DOL guidance specified that instead of enrolling all RESEA-eligible UI claimants in RESEA, states could enroll UI claimants with the highest *profiling score*—i.e., a statistical prediction of likely benefit exhaustion. Though this requirement no longer applies, some states still propose to enroll only claimants “most likely to exhaust” benefits. Instead, a state can consider enrolling all RESEA-eligible UI claimants in the evaluation if its sample size would be too small otherwise.

Similarly, if a state is considering excluding certain subgroups (e.g., veterans, claimants with disabilities) from the evaluation, it can examine its motivations for excluding them and determine whether exclusion is really needed. Sometimes these types of exclusions are not necessary.

The Bottom Line: Include in your impact evaluation all RESEA-eligible claimants, in as many offices as possible, for as long as is feasible.

2. SELECT A RESEARCH DESIGN TO REDUCE THE APPROPRIATE SAMPLE SIZE

RESEA programs can consider the following two evaluation design decisions to detect impacts with a smaller sample of study participants.

Choose an experimental design that uses random assignment.

Experimental designs use a lottery-like process to randomly assign claimants into two groups: a *treatment* group that receives the intervention and a *control* group that does not. Compared to non-experimental designs, experimental designs require smaller appropriate sample sizes (the total of treatment plus control). Because sample size is a particular challenge for smaller states, they can almost always select an experimental design (for more on experimental versus non-experimental designs, see the [RESEA Toolkit](#)).

Ensure that treatment and control groups are as close to the same size as possible.

If a state uses an experimental design, a critical choice to consider is the fraction of RESEA-eligible UI claimants to assign to each of the two experimental groups. As the sizes of the two groups become more unbalanced (in either direction), the appropriate sample size needed to detect impacts increases. Thus, the appropriate sample size is smallest when the ratio of claimants assigned to the treatment group and control group is one to one. In other words, the closer in number the treatment and control groups are to each other, the fewer people overall the state needs to enroll into the study.

Especially for evaluations of whole RESEA programs (compared to evaluations of program components; see Box 3), programmatic considerations will sometimes make it hard to enroll equal numbers of claimants in each study group. For example, many states have RESEA budgets that allow them to serve all or nearly all RESEA-eligible UI claimants. For a whole-program evaluation, having control and treatment groups of nearly equal size implies that some UI claimants the state has budget to serve will not be served and some RESEA dollars will be unspent. Thus, there is a trade-off between serving more claimants and a successful impact evaluation.

Box 3: Two Types of Impact Evaluations

1. **Whole-program evaluations** compare outcomes for those selected for RESEA versus outcomes for those not selected for RESEA—holding all else equal. That is, they try to estimate the impact of the “whole program.”
2. **Component evaluations** compare outcomes with the current RESEA program model versus outcomes with an alternative program model that changes some “component” of the program, such as the number or length of meetings—holding all else equal. That is, they try to estimate the incremental impact of a program component.

The Bottom Line: Adopt an experimental design using random assignment with treatment and control groups that are as equally sized as possible.

3. CONDUCT AN EVALUATION WITH SMALLER APPROPRIATE SAMPLE SIZES

RESEA programs can consider the following three actions to increase the magnitude of the potential impact.

Consider a whole-program evaluation.

In general, a state can expect the impact of a RESEA program component to be smaller (and often much smaller) than the impact of the RESEA program as a whole. It follows that appropriate sample sizes are substantially larger for component evaluations than for whole-program evaluations (see Box 3 above). Component evaluations are an attractive option because they are informative about whether a change in program design will improve program outcomes. Whole-program evaluations do not necessarily provide as much information on specific strategies to improve outcomes, but they could be a better option than a component evaluation if sample sizes are a concern.

Choose a component that plausibly has larger impacts.

Some components plausibly have larger impacts than others. In general, more-intensive components are expected to have larger impacts (see Box 4). In anticipating which components might have larger impacts, a state can consider its RESEA theory of action—*how the RESEA program achieves impacts*. Careful thinking about how a component might shift a pathway to higher employment and earnings can help to identify components with larger/smaller likely impacts. The webinar [An Introduction to RESEA Component Evaluations](#) further describes theories of action for component evaluations.

Box 4: Examples of More-Intensive Versions of a Component

- When using an impact evaluation to explore multiple meetings, consider six meetings instead of only two. Similarly, avoid having RESEA staff decide who would benefit from another meeting; instead, require every claimant to attend multiple meetings.
- Instead of adding another 15 minutes to the current 15-minute meeting, consider extending the meeting duration by one or more hours.

Recognize that different outcomes require different sample sizes.

DOL guidelines encourage estimating the impact of RESEA interventions on two outcomes: UI benefit duration and Q2 employment. As discussed, research on RESEA's predecessor program, REA, suggests that RESEA programs are likely to have relatively modest impacts on UI benefit duration and Q2 employment outcomes, and thus require large sample sizes to detect (Klerman et al., 2019). However, appropriate sample sizes vary with the outcome of interest. Appropriate sample sizes are largest for detecting impacts on employment and earnings, but smaller for detecting impacts on UI benefit duration.

Remember that every evaluation has value and can provide important information about state RESEA programs. This means that each state can estimate the impact of UI benefit duration and Q2 employment, even with a smaller-than-appropriate sample size. For some smaller states, a useful strategy can be to consider that their sample is likely large enough to detect the impact on UI benefit duration but not large enough to detect impact on Q2 employment—and make an effort to achieve the appropriate sample for at least UI benefit duration.

Box 5: Example of an Intermediate Outcome Evaluation

Michigan studied behavioral messaging to improve REA meeting attendance. The evaluation randomly assigned some selected UI claimants to more intensive and behaviorally informed meeting reminders; other selected UI claimants received only the conventional reminders (Darling et al., 2017).

The sample size was large enough to show that the new notices and reminders substantially improved meeting attendance, however, no impact on UI duration was detected. While the program logic model implies that meeting attendance leads to reduced UI duration, given the limited sample size, it cannot be confirmed.

Although DOL guidelines encourage estimating the impact of UI benefit duration and Q2 employment, states are not limited to those outcomes. When sample sizes are small, states can add other outcomes, including “intermediate outcomes” to their evaluations. Intermediate outcomes, such as meeting attendance, can also provide helpful information about a program (see Box 5) and require smaller sample sizes. As such, it is beneficial to identify intermediate outcomes with smaller appropriate sample sizes and include them in an evaluation, thereby establishing whether a programmatic change actually improves the intermediate outcome. The downside of intermediate outcome impact evaluations is that they cannot establish whether a programmatic change improves UI duration or Q2 employment. Again, it is important to consider evaluation goals and trade-offs.

The Bottom Line: Choose an intensive intervention that is expected to produce a larger impact or conduct a whole program evaluation. Carefully consider and select outcomes that require smaller sample sizes, including intermediate outcomes.

4. COMBINE EVIDENCE FROM MULTIPLE STATES

The final strategy involves a small state combining its evidence with evidence from evaluations in other states to create a stronger evidence base. There are two possible approaches.

Evaluate an intervention that other states are evaluating.

When a state does not have a sufficient sample of UI claimants for an impact evaluation, one option is to pool its sample with those from other states evaluating the same intervention to produce an adequate sample size. As long as several states are evaluating the same, or similar, interventions, the results from several states’ evaluations can be combined to produce a larger sample that can more reliably detect impacts. Abt Associates’ Evaluation Technical Assistance Team can help states with similar interventions to pool their data. The assistance might include describing what interventions are being evaluated in other states and providing information about how to coordinate the similar programs.

Even if your state is not explicitly partnering with another state, your results can be pooled with those of another state if you provide public-use or restricted-use analysis files with data from your evaluation that DOL or some other entity can combine with data from other states.

Alternatively, if your state reports enough detail about your intervention and your evaluation findings, DOL or a third-party evaluator could conduct a *meta-analysis* that systematically merges the findings of multiple single-state studies using statistical methods to calculate an overall impact. Meta-analysis does not simply pool data from smaller studies to achieve a larger sample size; rather, it uses statistical techniques so the overall estimate reflects differences across studies.

Use statistical methods to produce more precise estimates.

When estimates are statistically imprecise, as will often be the case for a small state with smaller than appropriate sample sizes, an evaluator can use statistical methods that augment the state's own findings with findings from other states. Rather than pooling across states, this kind of approach, based in what are known as “Bayesian” methods, still produces state-specific estimates. Those state-specific estimates shift (or, in technical parlance, “shrink”) the imprecise estimate for the state towards the average estimate of other states (Efron & Morris, 1973). To apply this approach, the evaluator would produce estimates for one state and produce a second estimate that combines it with estimates from other states. These Bayesian shrinkage methods will often detect an impact from a similar intervention in other states, even when no impact in the original state is detected. To use this approach, a state would need an evaluator with specialized expertise in Bayesian methods.

The Bottom Line: Combining results with other states or using statistical measures to improve estimates can reduce the need to gather a large sample from a single state.

SUMMARY

This brief presented four strategies that smaller states can use to align research goals with available samples. Often several of the suggested strategies can be used in combination to address the sample size challenge that small states face. If you are in a small state, consider this challenge and the most applicable solutions as early as possible in your evaluation planning; once you have secured an evaluator, continue to strategize with them.

The Bottom Line: Although achieving an appropriate sample size is important, don't let the “small state challenge” dissuade you from running an impact evaluation. All evaluations have value! With advance planning, a good evaluator, and support from DOL's evaluation technical assistance team, small states can conduct a rigorous and informative impact evaluation.

REFERENCES AND ADDITIONAL RESOURCES

Darling, M., O’Leary, C., Perez-Johnson, I., Lefkowitz, J., Kline, K., Damerow, B., Eberts, R., Amin, S., & Chojnacki, G. (2017). *Using behavioral insights to improve take-up of a reemployment program: Trial design and findings*. Mathematica Policy Research.

<https://www.dol.gov/sites/dolgov/files/OASP/legacy/files/13-50291-UIREA-FinalTechnicalReport-20170517.pdf>

Efron, B., & Morris, C. (1973). Stein’s estimation rule and its competitors—An empirical Bayes approach. *Journal of the American Statistical Association*. 68(341), 117–130.

<http://doi.org/10.2307/2284155>. JSTOR 2284155

Epstein, D., & Klerman, J. A. (2012). When is a program ready for rigorous impact evaluation? The role of a falsifiable logic model. *Evaluation Review*. 36(5), 375–401.

Klerman, J. A., Saunders, C., Dastrup, E., Epstein, Z., Walton, D., & Adam, T., with Barnow, B. S. (2019). *Evaluation of impacts of the Reemployment and Eligibility Assessment (REA) program: Final report*. Prepared for the U.S. Department of Labor. Abt Associates.

<https://www.dol.gov/sites/dolgov/files/OASP/evaluation/pdf/REA%20Impact%20Study%20-%20Final%20Report.pdf>

Klerman, J. A., Nightingale, D., Clarkwest, A., & Epstein, Z. (2022). *Options for building evidence on RESEA programs. Evaluation to advance RESEA program evidence*. U.S. Department of Labor.

https://www.dol.gov/sites/dolgov/files/OASP/evaluation/pdf/resea/Options_for_Building_Evidence_on_RESEA_Programs_final_508c.pdf

Mills De La Rosa, S., Souvanna, P., Clarkwest, A., Kappil, T., Epstein, Z., Rothschild, L., Kuehn, D., Wall, A., Klerman, J., & Nightingale, D. (2021). *Reemployment Services and Eligibility Assessment (RESEA) evaluation toolkit: Key elements for state RESEA programs*. U.S. Department of Labor.

https://www.dol.gov/sites/dolgov/files/OASP/evaluation/pdf/RESEA_Toolkit_February2021.pdf

Trutko, J., Trutko, A., Clarkwest, A., Souvanna, P., Klerman, J. A., Spaulding, S., Briggs, A., Scott, M., Hecker, I., Islam, A., & Katz, B. (2022). *RESEA program strategies: State and local implementation*. Report submitted to U.S. Department of Labor, Chief Evaluation Office. Abt Associates.

https://wdr.doleta.gov/research/FullText_Documents/ETAOP2022-17_RESEA%20Program_Strategies_State_and_Local_Implementation_Final_Report.pdf

U.S. Department of Labor. (2019). *Expectations for states implementing the Reemployment Services and Eligibility Assessment (RESEA) program requirements for conducting evaluations and building program evidence*. Unemployment Insurance Program Letter No. 1-20. Employment and Training Administration, U.S. Department of Labor. https://wdr.doleta.gov/directives/attach/UIPL/UIPL_1-20.pdf

Suggested citation: Klerman, J. A., & Wheat, V. (2023). *Strategies for RESEA evaluation in small states*. RESEA Evaluation Technical Assistance Brief. Prepared for the U.S. Department of Labor, Chief Evaluation Office. Abt Associates.