

Frequently Asked Questions about RESEA Evaluations

November 2020

Contents

General Questions	2
Impact Evaluations.....	4
Non-Impact Evaluations	8
Sample Sizes	9
Logistics of Implementing an Evaluation	11
Pooled Evaluations	12
Protecting Human Subjects of Research.....	13

About this FAQ

This RESEA Evaluation FAQ contains responses to frequently asked evaluation related questions of a *technical nature*.

For additional technical assistance (TA) resources on RESEA evaluations, please check out the RESEA Evaluation and Evidence Resources page on WorkforceGPS at:

https://rc.workforcegps.org/resources/2019/07/30/17/32/RESEA_Evaluation_Evidence_Resources.

Resources will be posted there as they become available. Additional evaluation resources and reports may be found at the Chief Evaluation Office's website at: <https://www.dol.gov/agencies/oasp/evaluation>.

If you need additional help or have specific questions, please contact DOL's RESEA evaluation TA Help Line at RESEA@abtassoc.com for further assistance.



General Questions

What is an evaluation? How is an evaluation different from other types of evidence-building activities, such as performance management?

As part of program management and operations, states generate and report data for performance accountability. Performance accountability is an evidence-building management function that involves ongoing reporting of program activities and progress towards planned outcome goals. Evaluation, on the other hand, is an evidence-building function that uses “systematic data collection and analysis of one or more programs, policies, and organizations intended to assess their effectiveness and efficiency.”¹ This includes focusing more on questions of “how” and “why.” For example, as the U.S. Office of Management and Budget describes:

Evaluations may address questions related to the implementation of a program, policy, or organization; the effectiveness of specific strategies related to or used by a program, policy, or organization; and/or factors that relate to variability in the effectiveness of a program, policy, or organization or strategies of these. Evaluations can also examine questions related to understanding the contextual factors surrounding a program, as well as how to effectively target specific populations or groups for a particular intervention. (U.S. OMB 2019, p.200-16)²

Evaluations often also differ in *who* is involved in conducting them. Performance reporting is typically performed solely by the relevant government agency. Evaluations usually involve an independent evaluator—perhaps in a different state office or agency, perhaps outside state government—who can provide an independent perspective and specialized skills in the methods required for the particular type of evaluation.

What are the different types of evaluations?

Different types of evaluations employ different methods and designs to respond to research questions. The following types of evaluations are the most common to inform policy development and ongoing program improvement:

- **Impact Evaluation:** This type of evaluation assesses the impact of a program or component of a program on outcomes, typically relative to a counterfactual situation. An impact evaluation uses statistical methods to estimate what happens in absence of the program or component of the program. Impact evaluation includes both experimental (i.e., randomized controlled trials) and quasi-experimental designs. These types of evaluations speak to the “does it work?” question.
- **Outcome Evaluation:** This type of evaluation measures the extent to which a program has achieved its intended outcome(s), and focuses on outputs and outcomes to assess effectiveness. Unlike an impact evaluation, an outcome evaluation cannot show causal

¹ Foundations for Evidence-Based Policymaking Act of 2018, citing 5 U.S. Code § 311(3).

² Office of Management and Budget. *Circular No. A-11: Preparation, Submission, and Execution of the Budget*, June 2019. URL: <https://www.whitehouse.gov/wp-content/uploads/2018/06/a11.pdf>.

impacts. An outcome evaluation can help answer questions like, “Did the program, policy, or organization do what it intended to do?”

- **Implementation or Process Evaluation:** These types of evaluations assess how the program or service is delivered relative to its intended design and theory of change (basically the thinking behind how your intervention will achieve results),³ and often include information on content, quantity, quality, and structure of services provided. Implementation or process evaluations can be conducted on their own, but are often conducted along with impact and/or outcomes evaluations. Implementation or process evaluations can help answer questions like, “Was the program or policy implemented as intended?” or “How is the program, policy, or organization operating in practice?”
- **Formative Evaluation:** This type of evaluation, typically done before full implementation of a program, assesses whether a program or component of a program is feasible, appropriate, and acceptable before it is fully implemented. It may include some of the activities described above, such as process evaluation or outcome evaluation. However, unlike summative evaluation designs like impact and outcome evaluations, which seek to answer whether or not the program met its intended goal(s) or had the intended impact(s), a formative evaluation focuses on refining program operation, given the program model. A summative evaluation can then explore whether a well operating program model actually leads to the desired outcomes.

What is DOL’s Clearinghouse for Labor Evaluation and Research (CLEAR)? How might it help my state design or evaluate its RESEA program?

DOL established the Clearinghouse for Labor Evaluation and Research (CLEAR)⁴ to make research on labor topics more accessible to practitioners, policymakers, researchers, and the public so that evidence can inform policy and program decisions. To achieve this goal, CLEAR conducts systematic evidence reviews of research and evaluation reports on labor topics, and then summarizes those studies on the CLEAR web site.

As part of its systematic reviews of research, CLEAR assigns causal evidence ratings to impact *studies*. This rating is an indicator of the credibility of a study’s findings, or the level of confidence you can have that the study’s findings truly reflect the causal impact of the intervention studied (or lack thereof) and not some other factor.⁵ CLEAR also has guidelines for high quality quantitative descriptive and implementation studies, but does not currently assign evidence ratings to those types of studies. The causal evidence ratings CLEAR gives to impact studies are represented by the gas gauge icon. Those ratings do not change after they are assigned.

³ A theory of change is a way to explain how your intervention’s design produces its intended outcomes. The webinar [“Evaluating RESEA: How Does it Help My State and Where Do We Start?”](#), available on [Workforce GPS](#), describes the value of making sure that staff have a shared understanding of an intervention’s theory of change and how you can create a visual representation of an intervention’s theory of change.

⁴ CLEAR may be found here: <https://clear.dol.gov>.

⁵ A high quality study can provide strong evidence that a program does not work; put differently that the program being evaluated is low quality.

Icons indicating CLEAR's rating of a study's credibility of evidence



CLEAR also assigns causal effectiveness ratings to RESEA *interventions* studied in new and ongoing evaluations. The rating is based on findings from the entire body of credible impact studies⁶ of that intervention, taken together. Causal evidence ratings are represented by the thermometer icon.

Icons indicating CLEAR's rating of an interventions' effectiveness, from all studies of the intervention



These ratings of evidence of intervention effectiveness *may change over time*, as more evidence becomes available (including evidence from impact evaluations conducted by states). As more evidence on RESEA programs becomes available, CLEAR will conduct periodic evidence reviews to identify new studies in this evidence base and assign causal evidence ratings based on both study quality and effectiveness of the intervention examined in the study. For more information about CLEAR's processes and how impact studies and interventions are rated, visit CLEAR's reference documents.⁷

Impact Evaluations

What types of impact studies can meet CLEAR standards?

CLEAR rates impact studies that use experimental designs and a range of quasi-experimental designs. Two types of designs are currently eligible to earn a High rating from CLEAR. Those designs are:

- (1) well-conducted **randomized control trials** (RCTs) that have low attrition and no other threats to study validity and
- (2) **Interrupted time series** (ITS) study designs with sufficient replication

Other types of quasi-experimental designs that are not eligible for a High rating can earn a Moderate rating. Of particular note are matched comparison group designs, which can receive a Moderate rating if the study's comparison group is demonstrated to be sufficiently similar to the

⁶ That is, all impact studies that earned a Moderate or High rating from CLEAR.

⁷ Find CLEAR's reference documents here: https://clear.dol.gov/about?qt-about_clear_tabs=3#qt-about_clear_tabs.

study's intervention group. Studies using RCT and ITS designs that fail to qualify for a High rating may otherwise be able to earn a Moderate rating. Any studies that don't qualify for a Moderate rating receive a Low rating.

For more detailed information on the standards used to rate an impact study's quality of evidence, please refer to the *CLEAR Causal Evidence Guidelines, Version 2.1*. That document can be found on CLEAR's website at:

https://clear.dol.gov/sites/default/files/CLEAR_EvidenceGuidelines_V2.1.pdf.

CLEAR does not currently have standards for reviewing studies that use regression discontinuity (RD) designs. Should that prevent a state from considering RD as a quasi-experimental impact evaluation design option?

Knowing that many states use profiling models to select RESEA claimants, states may naturally be interested in regression discontinuity (RD) as an impact evaluation design option. Although CLEAR has not yet established standards for reviewing and rating the quality of RD designs, RD is a design that can produce strong causal evidence, and states and their independent evaluators may give RD designs full consideration as an option for RESEA evaluations. If a state believes that RD is the most rigorous and appropriate design for it to use in evaluating RESEA, states and their independent evaluators may refer to the standards that the U.S. Department of Education's What Works Clearinghouse currently uses to review RD studies in designing and conducting their evaluations.⁸

What can states do to help ensure that their impact evaluations can meet CLEAR's standards for the credibility of study findings?

CLEAR reviews impact studies as well as other types of studies. But only impact studies receive ratings for the credibility of study findings. For impact evaluations, obtaining CLEAR's High or Moderate rating for the study's credibility is the first step toward an intervention being able to be rated as High or Moderate for its evidence of effectiveness.

There are three related steps states can take to help their RESEA impact evaluations meet CLEAR's causal evidence standards⁹ to improve the quality of evidence produced:

- 1) Use an **academically rigorous design with comparison groups**, like random assignment;
- 2) Have a **large enough sample**, so the results are statistically significant (see further discussion on sample sizes below); and

⁸ Please refer to the "What Works Clearinghouse Standards Handbook, Version 4.0." URL: https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_standards_handbook_v4.pdf.

⁹ CLEAR's standards for rating the quality of evidence produced by individual studies can be found here: https://clear.dol.gov/sites/default/files/CLEAR_EvidenceGuidelines_V2.1.pdf.

- 3) Work with an **experienced, independent evaluator** with no direct or indirect responsibility for program administration or operations. An experienced evaluator can help a state select and implement an appropriate impact evaluation design.¹⁰

For an impact study's results to contribute to a High or Moderate rating of *intervention* effectiveness, that study must not only satisfy CLEAR's standards for *study* quality, but must also find statistically significant beneficial impacts on employment and UI duration¹¹. The smaller a study's sample size, the greater the risk that it will not find statistically significant impacts, even if the intervention being evaluated is effective. An experienced evaluator can also help identify how large of a sample is appropriate for your evaluation.

Does random assignment mean that states cannot target RESEA claimants based on profiling scores or other non-random factors?

States may be concerned that using random assignment in their RESEA evaluation will conflict with using RESEA to meet Worker Profiling and Reemployment Services (WPRS) requirements, which include targeting claimants based on profiling scores. Researchers have successfully used random assignment to evaluate RESEA's predecessor, REA, in several states.¹² DOL can work with individual states to explore potential flexibilities that allow states to conduct an evaluation while they maintain compliance with the WPRS requirements.

Examples of such flexibilities include:

- *Evaluating a program component.* If you plan to evaluate a component of your program, then there is likely no issue with WPRS requirements. You can select RESEA claimants based on their risk of UI benefit exhaustion as you would otherwise. Then, after claimants have been selected, some are randomly assigned to receive RESEA with the component and others are randomly assigned to receive the program without that component.
- *Using a multi-faceted RESEA selection process.* For evaluations of the whole program there may be opportunities to build random assignment into the WPRS profiling model, or apply randomization to a non-WPRS segment of the RESEA population.

How might random assignment be performed?

Generally, random assignment is automated. For example, code can be written into the system that states use to select program participants—in this case, the system the state uses to select

¹⁰ See also the RESEA evaluation TA webinar, "[Procuring and Selecting an Independent Evaluator](#)" on [WorkforceGPS](#) for further information on how to assess an evaluator's experience and qualifications.

¹¹ See Unemployment Insurance Program Letter No. 1-20 for more details on the criteria for RESEA intervention causal evidence ratings at: https://wdr.doleta.gov/directives/attach/UIPL/UIPL_1-20_acc.pdf

¹² Klerman, J. A., Saunders, C., Dastrup, E., Epstein, Z., Walton, D., and 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. Cambridge, MA: Abt Associates.; Poe-Yamagata, E., Benus, J., Bill, N., Carrington, H., Michaelides, M., & Shen, T. (2011). *Impact of the Reemployment and Eligibility Assessment Initiative*. Prepared for the U.S. Department of Labor. Columbia, MD: IMPAQ International.

UI claimants for RESEA and to send scheduling letters. The system will then automatically assign eligible participants to “treatment” and “control” groups. Manual approaches (i.e., approaches that are not inserted into the computer system used for selection) at the state level are also possible. For example, a spreadsheet could be created for staff to use to perform random assignment. However, when randomizing thousands of cases, such manual systems are labor intensive. Also, regardless of the number of cases randomized, manual processes are more likely to introduce errors into the assignment process.

Whether an automated or manual system is chosen for randomization, the state and its evaluator must regularly monitor and check the results of the assignment process, to ensure that the distribution of those assigned is consistent with the intended process. If intervention and control claimants differ (e.g., in their average prior earnings, age distribution, where in the state they live, etc.), it would indicate that the assignment mechanism is not functioning as intended. In general, those checks are automated, to make them less labor intensive. Work with your experienced evaluator and the staff that manage data and IT systems in your state to determine the best way forward for your evaluation.

For purposes of a random assignment impact evaluation, is it a problem to allow the control group members to receive similar services from AJC staff?

Typically, no. Impact analyses intend to examine impacts of the RESEA intervention versus what would have occurred had the claimant not been selected for the intervention. Under normal circumstances (i.e., in the absence of the evaluation), non-RESEA claimants have access to AJC services. In a random assignment impact evaluation of an RESEA as a whole, it is therefore also appropriate for control group claimants to be able to seek out and receive AJC services. For the purposes of the evaluation, the control group receives the “business as usual” condition; in other words, the state UI agency should only point those assigned to the control group to AJC services to the same extent that it would for non-RESEA claimants in the absence of the evaluation.

When thinking about interventions to evaluate, states and their evaluators should keep in mind the types of services to which the comparison group may have access. For the purposes of statistical power, it is best that the intervention the treatment group receives be as different as possible from those services available to the control group. The less distinct the services are, the smaller the expected impact will be. Furthermore, small impacts imply that a larger sample would be required to detect the likely impacts.¹³

What qualifies as a valid comparison group for a quasi-experimental design?

A valid comparison is one that is, on average, like the intervention group in all ways other than their access to the intervention. If the groups differ in ways that may affect outcomes—e.g., they differ on prior work history, educational attainment, local labor market conditions, motivation,

¹³ See also the RESEA evaluation TA webinar, “[De-Mystifying Random Assignment](#)”, on [WorkforceGPS](#) for further information on random assignment evaluations.

etc.—then subsequent differences in outcomes between the groups might be the result of those pre-existing differences between claimants, rather than being the result of the intervention.

Properly implemented, random assignment always delivers a valid control group. In the absence of random assignment, generating a valid comparison group is challenging. In general, those not selected for RESEA differ from those selected in some specific way. Most particularly, they tend to have lower profiling scores—so in the absence of RESEA services, we expect that non-RESEA claimants would be more likely to find a job before they exhaust their UI benefits. The fact that RESEA and non-RESEA claimants differ systematically from one another in such fundamental ways will usually make non-RESEA claimants an invalid comparison group.¹⁴

Please refer to “Criterion Regression.1 Were the intervention and comparison groups similar before the intervention?” in CLEAR’s *Causal Evidence Guidelines, Version 2.1*¹⁵ for a more detailed description of what constitutes a valid comparison group.¹⁶

Non-Impact Evaluations

What if my state is not in a position to start an impact evaluation now?

One of the goals of the new RESEA program is to ensure that each state employs RESEA interventions and service delivery strategies that are based on rigorous, credible causal evidence¹⁷ and are shown to reduce UI benefit duration and as a result of improved employment outcomes. The requirement that RESEA funds be used only for interventions demonstrated to reduce the number of weeks for which program participants receive unemployment compensation by improving employment outcomes is already in effect and this requirement will become much more rigorous beginning in FY 2023 when states must be able to link funds to evidence-based models and interventions. The most effective way to accomplish this goal is for states to begin evaluating the impact of your RESEA interventions and service delivery strategies as soon as possible.

Nevertheless, in some cases a state might not be ready to conduct an impact evaluation yet. For example, a state might not be sure that its data systems can support the needs of an evaluation. Or, it might not be sure that new interventions are currently being implemented as intended. Or, it may not be sure what intervention would be useful to study.

In such cases, it may make sense for a state to start by conducting evaluation activities or other types of studies (discussed earlier) that build evidence that lead to an impact evaluation, such as:

- Implementation or process studies: An implementation study can help you better understand how RESEA programs are being implemented across the state, and if

¹⁴ See also the RESEA evaluation TA webinar, “[What Evaluation Designs Are Right for My State?](#)”, on [WorkforceGPS](#) for further information on potential comparison groups and establishing a credible counterfactual.

¹⁵ URL: https://clear.dol.gov/sites/default/files/CLEAR_EvidenceGuidelines_V2.1.pdf.

¹⁶ See also the RESEA evaluation TA webinar, “[Evaluations Using Your Existing Administrative Data: Quasi-Experimental Designs](#),” on [WorkforceGPS](#) for further information on quasi-experimental comparison groups.

¹⁷ That is, evidence from studies that receive a High or Moderate rating.

implementation matches the intended design. A client flow study can tell you how frequently different types of services are being used by RESEA claimants. Each of those may help identify problems to fix or promising components to evaluate.¹⁸

- **Outcome studies:** An analysis of outcomes, including how they vary across subgroups of claimants may also help identify places where improvements might be needed before beginning an impact study.
- **Evaluability assessment:**¹⁹ This is an assessment to determine your state's readiness for an impact evaluation. It may include implementation or process studies to better understand how your program is being implemented, the quality of your data, and your data's ability to fully track claimants' activities and outcomes. Other activities to verify the quality, completeness, and availability of data may also be helpful.²⁰

Where can I find examples of good implementation or process studies?

CLEAR's website includes guidelines for quality implementation²¹ and descriptive²² types of studies. Additionally, some CLEAR topic areas do include implementation, process, and outcome studies as well as impact evaluations (though the reemployment topic area does not). Although CLEAR does not rate those studies on the strength of their findings, they do provide useful information about how the program is designed, aggregate participation and outcome levels, and organizational information.

Sample Sizes

How large a sample will my state's impact evaluation need to have?

There are no hard-and-fast thresholds for exactly how large a sample a study is appropriate. The smaller the sample size, the greater the risk that the study could fail to find statistically significant impacts, even if the intervention is effective. Although there are no exact thresholds for sample size, what you will want to make sure of is that the sample size that your state is able to generate is as large as possible and is within the "ballpark" of what's needed to obtain statistically significant results. For random assignment studies, that ballpark is rarely less than a few thousand, often more than 10,000. For certain questions sample sizes of more than 100,000 might be needed in order to have a good chance of being able to find impacts. Based on specifics of your state's research question and design, experts such as research units from

¹⁸ Please view the RESEA evaluation TA webinar, "[Implementation Studies](#)," on [WorkforceGPS](#) for further details.

¹⁹ Please view the RESEA evaluation TA webinar, "[Evaluating RESEA: How Does it Help My State and Where Do We Start?](#)" on [WorkforceGPS](#) for further details on evaluability assessments.

²⁰ Please view the RESEA evaluation TA webinar, "[Assessing Data for Evaluation](#)," on [WorkforceGPS](#) for further details.

²¹ CLEAR's guidelines for reviewing implementation studies can be found here: <https://clear.dol.gov/reference-documents/clearinghouse-labor-evaluation-and-research-clear-guidelines-reviewing-0>.

²² CLEAR's guidelines for reviewing quantitative descriptive studies can be found here: <https://clear.dol.gov/reference-documents/quantitative-descriptive-guidelines>.

within your agency or an independent evaluator can help you make sure that you are, in fact, in the sample size “ballpark.”

The “ballpark” for the number of claimants a study needs to include to generate statistically significant impacts depends on several considerations. The most fundamental considerations are:

1. What intervention are you evaluating? Is it your RESEA program as a whole, or some component of it?
2. What outcomes are of interest?

Sample sizes for an evaluation will also depend on the type of impact design that a study uses.²³ Random assignment studies have the smallest sample size requirements. Quasi-experimental designs (QEDs) vary substantially in the sample sizes required to detect impacts, but are often several times larger than for random assignment.

Exhibit 1 presents “ballpark” estimates, informed by prior studies, of the sample sizes likely to be required for an impact evaluation that uses random assignment to estimate impacts. It presents estimates for various combinations of interventions and outcomes.

EXHIBIT 1. BALLPARK ESTIMATES OF APPROPRIATE SAMPLE SIZES FOR RANDOM ASSIGNMENT EVALUATIONS

<i>Intervention</i>	<i>“Small” Component</i>	1,000-3,000	100,000+	Hundreds of Thousands
	<i>“Large” Component</i>	500-1,000	50,000-100,000	100,000+
	<i>Whole Program</i>		5,000-10,000	10,000-25,000
		<i>Meeting Attendance</i>	<i>UI Duration</i>	<i>Employment</i>
		<i>Outcome</i>		

Examples of “smaller” RESEA service components that might be evaluated include labor market information or introducing self-scheduling. A “larger” component might be intensive case management or a reemployment services package as a whole.

An experienced evaluator can conduct a type of analysis called “power analysis,” which can give you a more informed estimate of how many sample members your study would need, based on the study’s design, the intervention that you are evaluating, and the outcomes that you are examining.

Although your state may not serve enough claimants to generate the sample sizes needed to answer *every* question of potential interest, no matter how small your state is there are important research questions that your state *does* have the sample size to address.

²³ See also the RESEA evaluation TA webinars, [“De-Mystifying Random Assignment”](#) and [“Evaluations Using Your Existing Administrative Data: Quasi-Experimental Designs,”](#) on [WorkforceGPS](#) for further information on the sample sizes needed for different evaluation designs.

What if my state does not serve enough claimants in a year to achieve the sample size to support the kind of impact evaluation my state agency has in mind?

Two main types of options exist:

- The first is to find ways to increase the available sample size.
- The second is to change your study in ways that would reduce the sample size required.

One straightforward way to get a larger sample is to collect more years of data. For example, randomize claimants for two or more years, rather than one. Another possibility is to join with other states in a joint evaluation studying the same or very similar intervention and pool your samples.

Regarding ways to reduce how large a sample your state needs, one option is to reconsider the intervention being evaluated. Choosing a larger component or whole program, rather than a smaller component, would reduce the sample size required. Another approach to reduce the sample size required is to choose an evaluation design that does not require as large of a sample. Random assignment designs have the smallest sample size requirements.

Logistics of Implementing an Evaluation

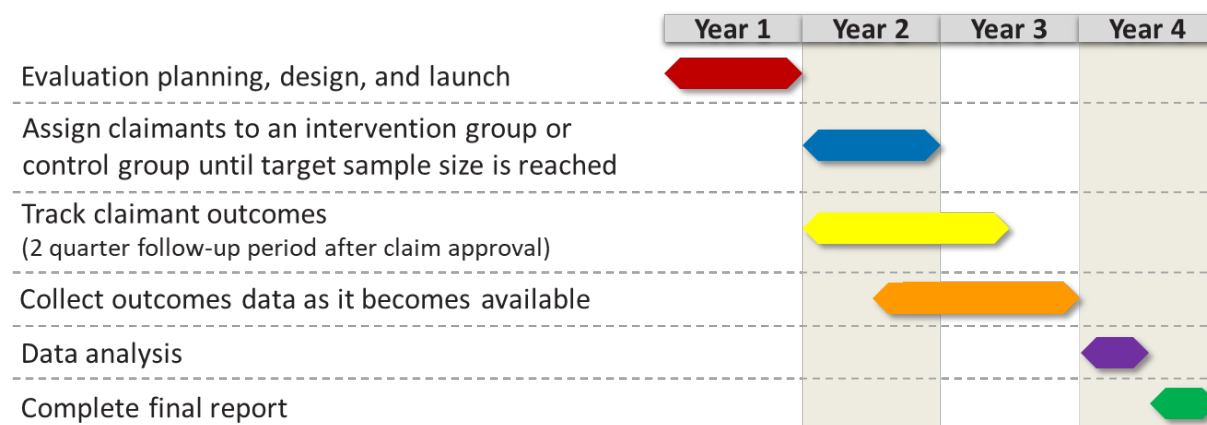
What kind of information technology (IT) support is necessary to conduct an evaluation?

Some level of IT support will typically be necessary to conduct an evaluation, though that varies substantially by type of evaluation. An implementation or process study will require fewer IT resources. As a rough estimate, for impact studies, states will likely need multiple weeks of IT staff time before the evaluation launch (to develop, insert and test random assignment algorithms and make changes to data systems) and then about a day per quarter during the evaluation and follow-up period for data extraction.

How long will it take to complete an evaluation?

Impact evaluation timelines are determined by whether the study is prospective or retrospective. A *prospective* study enrolls claimants and measures their outcomes after the evaluation has started. Random assignment evaluations are prospective. Some quasi-experimental evaluations may be prospective as well. Exhibit 2 shows rough timelines for activities involved in a prospective study. Expect around a year for planning, a year (or more) to enroll UI claimants, a year for outcomes to occur and data to become available, and a year for analysis, write-up, and review.

EXHIBIT 2. ROUGH TIMELINES FOR A PROSPECTIVE IMPACT EVALUATION



A *retrospective* impact study uses existing data on characteristics and outcomes of claimants served in the past. For RESEA evaluations, this will typically be administrative data. Because it uses existing data on past participants, a retrospective study can be completed more quickly. Planning and launching a retrospective impact study is somewhat less complicated than planning and launching prospective studies because no programmatic changes are made, and time is not required to assign claimants and track outcomes. On the other hand, because retrospective studies use QED methods, and because QED analysis is more complicated, the analysis phase may take longer. But all-in-all, the total time for a retrospective study is notably shorter than for a prospective study.²⁴

Pooled Evaluations

What should states weigh when considering whether to join with other states in a pooled evaluation?

Participating in a pooled evaluation with other states offers some important potential benefits. For example it allows for a larger sample, which allows your evaluation to examine questions that might not be feasible to evaluate otherwise. In particular, pooled evaluations might make impact studies feasible for “smaller” states (those serving less than 15,000 claimants annually). Pooled studies might also permit evaluation of RESEA components of interest, not just a state’s RESEA program as a whole. A pooled evaluation lowers the cost for each state.

But conducting a joint evaluation also entails some logistical considerations. Below are some questions to consider:

- What is your state’s willingness and ability to share control over an evaluation with other states? In a pooled evaluation, more decisions will need to be made by consensus with your partners. This includes coordinating decisions over the design of the intervention to be tested (since that intervention must be similar across states), how to select an evaluator, and how to oversee that evaluator’s work.

²⁴ See the RESEA evaluation TA webinar, [“What Evaluation Details Do I Need for a Plan and How Long will it Take?”](#) on [WorkforceGPS](#) for more detail on thinking through evaluation timelines.

- What state-specific constraints might exist in your state that would hinder contracting with other states? Some states may have legal or other regulatory restrictions that limit or otherwise present challenges to their ability to enter into the agreements with other states to share needed data or to procure an independent evaluator from outside of their state.

Are pooled evaluations only valuable for “small” states?

No! The technical and financial benefits of participating in a pooled evaluation are relevant to states of any size. For example, to evaluate some interesting, but smaller, components requires samples larger than most any state can generate on its own.

Protecting Human Subjects of Research

Do states need to seek review by an Institutional Review Board (IRB) before beginning their evaluations?

In conducting evaluations, states should always ensure appropriate protection of the human subjects who participate in the study. It is important that determinations about whether a study adequately protects human subjects be made by a qualified independent entity that does not have a stake in the study. That type of entity is known as an Institutional Review Board (IRB). An IRB is “an appropriately constituted group that has been formally designated to review and monitor... research involving human subjects.”²⁵

IRBs review research designs that involve human subjects to ensure human subjects are protected and the informed consent process is appropriate and comprehensive. In some cases, the IRB may determine that your design does not pose a human subjects risk and is exempt from full review and/or is exempt from the requirement to provide informed consent. However, that determination can only be made by the IRB, and not by DOL, the state, or the state’s evaluator. IRB approval or an exemption is required before beginning an evaluation.

All universities, most research institutes, and some state agencies have IRBs. Whether or not your independent evaluator has its own IRB, they should know where to find and how to submit materials describing your evaluation to an IRB. When selecting your independent evaluator, one important consideration is that your evaluator has a strong plan for the protection of human subjects and getting IRB review of study plans and materials.²⁶

²⁵ From the U.S. Food and Drug Administration’s information sheet, “Institutional Review Boards Frequently Asked Questions Guidance for Institutional Review Boards and Clinical Investigators January 1998.” URL: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/institutional-review-boards-frequently-asked-questions>.

²⁶ See the RESEA evaluation TA webinar, “[What Evaluation Details Do I Need for a Plan and How Long will it Take?](#)” on [WorkforceGPS](#) for more detail on the IRB review process.