Evaluation of the Reemployment and Eligibility Assessment (REA) Initiative

How to Construct a Comparison Group

Technical Assistance Paper August 2008

Table of Contents

Evaluation of the Reemployment and Eligibility Assessment (REA) Initiative How to Construct a Comparison Group

Technical Assistance Paper
August 2008

Selected states implemented the Reemployment and Eligibility Assessment (REA) Initiative in 2005. These states were required in 2006 to incorporate into their REA Initiative a comparison group made up of individuals not receiving REA services. This comparison group was to be used in evaluating the effectiveness of the REA Initiative.

To date, states have experienced varying degrees of success in designing and implementing comparison groups to test the effectiveness of the REA Initiative. The purpose of this document is to set forth the requirements for a rigorous evaluation design. The paper begins with a discussion of why it is important to rigorously evaluate the REA Initiative. Next, the paper presents an overview of random assignment evaluations, a simple random assignment design, how to measure program impacts, and how to determine the sample size needed. The document concludes with a number of best practices that state managers should keep in mind when designing and implementing this type of evaluation design.

A. Why Measure REA Impacts?

To accurately measure the impact of new program initiatives like REA, policymakers need evidence-based results obtained from rigorous quantitative evaluations. Such evaluations are important because they help administrators and policymakers fulfill their responsibilities and achieve their goals. Evidence-based evaluation results have been used to:

- inform policy makers about new policies or initiatives,
- convince legislators to fund new initiatives,
- refine the design and targeting of existing initiatives,
- justify on-going budget requests, and

- cut spending on ineffective initiatives.

Particularly for new initiatives such as REA, policymakers need solid evidence on whether the program is achieving its goals. Solid evidence on the net impact of the program is important not only for DOL policymakers, but also for Office of Management and Budget analysts and Congressional decision makers. Congress often requires strong evidence derived from rigorous program evaluation to fund new initiatives like REA.

B. Evaluation Design Overview

The underlying objective of virtually all public benefits programs is to improve the lives or outcomes of program beneficiaries relative to what they would have been in the absence of the program. To evaluate the impact of the REA Initiative, researchers must be able to measure the difference between the outcomes actually experienced by beneficiaries (e.g., their duration of unemployment, earnings, etc.) and the outcomes that would have been achieved in the absence of the initiative.

It is difficult to measure the impact of an initiative because we cannot directly observe what would have happened to the initiative's beneficiaries in the absence of the initiative (i.e., the "counterfactual"). That is, because the program was implemented, we do not know the counterfactual - what would have happened in the absence of the program. One way of "observing" the counterfactual is to randomly assign potential beneficiaries to two groups:

- the "treatment" group which is invited to participate in program services or benefits, and
- the "comparison" group which is excluded from program services.[1]

"Random assignment" of potential beneficiaries to the treatment or the comparison group means assignment on the basis of a chance event, like the flip of a coin. In practice, a computerized random number generator is usually used to make the assignments. Because assignment is purely by chance, there can be no systematic differences between the two groups except exposure of one group to the treatment. In other words, the two groups (treatment and

[1] In experimental design evaluations, "comparison" group is sometimes called "control" group.

comparison) will be similar in their demographic characteristics, labor markets characteristics, as well as other variables which would affect economic outcomes. Thus, any observed differences in average outcomes between the two groups can be confidently attributed to exposure to REA services and not to selection bias.

While a random assignment design is the gold standard for measuring program impacts, it is not the only methodology available for measuring program impacts. One alternative to random assignment is the use of a quasi-experimental methodology (such as propensity score matching) for creating a comparison group and measuring program impacts. The key issue in quasi-experimental impact evaluations is the identification of an appropriate comparison group that is similar to the participant group.[2] Once an appropriate comparison group is identified, observed differences in outcomes between the treatment and comparison groups can be attributed to the program.

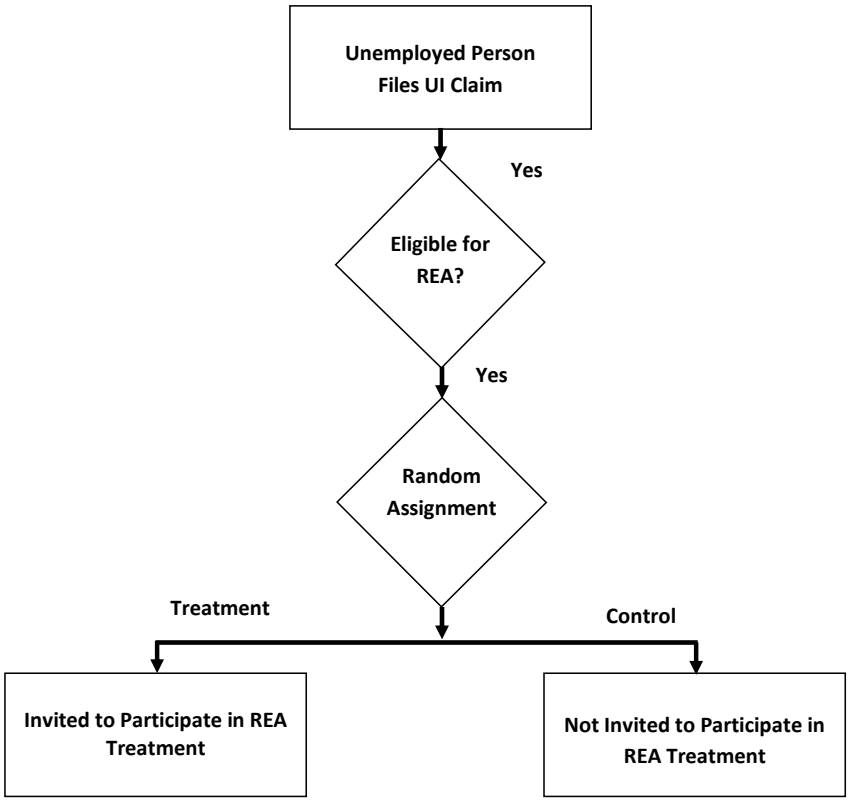
C. Simple Random Assignment Design

Figure 1 presents an effective random assignment design for measuring REA impacts. Some states have successfully implemented this design. This simple design has a number of advantages: it is straightforward to implement, it dramatically simplifies the measurement of program impacts, and it reduces the complexity of reporting data for the ETA 9129. This design includes only a few simple steps:

1) After claimants file their UI claims, they are assessed for eligibility for the REA Initiative based on the state's criteria for REA participation.
2) Claimants found to be eligible for REA are then randomly assigned into either the REA treatment group or the comparison group.
3) The REA treatment group receives REA services (or at least is invited to), while the comparison group does not. The most powerful design has groups of equal size, but if necessary, different proportions can be assigned to each group. (See comments later in the paper.)

[2] Attachment I provides additional information about quasi-experimental methodologies.

Figure 1: Random Assignment Design



Within this basic random assignment design, states have substantial discretion. For example, some states may wish to offer REA services early in an individual's UI claim while other states may want to provide REA services later in the claim. In practice, some states have chosen to begin REA services in the second week of a claim; others have chosen to initiate REA services in the fifth week and even later.

Whether the state decides to initiate REA services in the second week or the fifth week of a claim (or any other week), it is important that random assignment be implemented immediately prior to the receipt of REA services. For example, if REA services are to be provided in week five of a claim, random assignment should take place in week four. It is also critical that both treatment and comparison group members be randomly assigned at the same point in time. For example, it is not appropriate to assign the comparison group in week two and the treatment group in week four.

D. Evaluating Program Impacts

To evaluate program impacts, we measure the difference in average outcomes (e.g., reemployment) between the treatment and the comparison groups. The treatment group should include anyone who is invited to participate in an REA (e.g., those who receive the REA letter or invitation). The comparison group should include all those individuals who were eligible for REA selection, but were not selected. The comparison group members should not receive an REA invitation or other REA service at any point. Comparison group members, of course, may receive any other services that are available to all UI claimants.

E. Sample Size

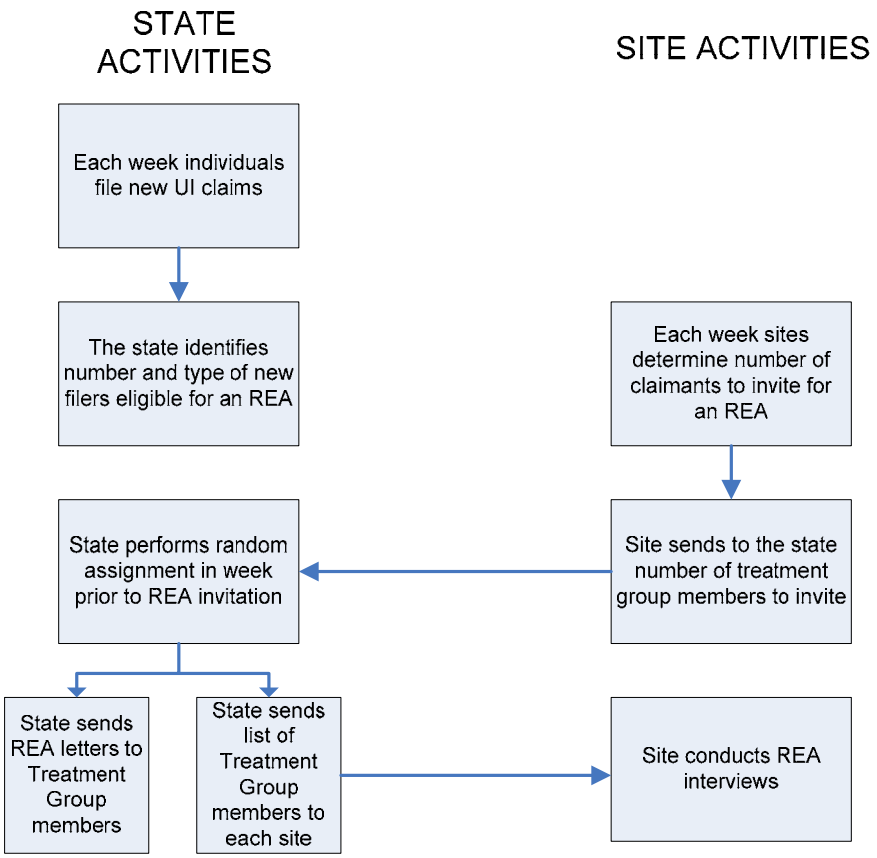
The sample size needed for a random assignment evaluation is determined by several factors: the size of the impact for treatment you want to be able to detect, the size of the standard deviation of the key variable in the population, the power with which you wish to be able to find an effect statistically significant if it exists (power), and the level at which you want to avoid finding a result from your sample by chance as opposed to a true effect (alpha level). In general, the more people involved in the study, the more power the study has to detect smaller effects. The most efficient design divides the total sample into equal-sized treatment and comparison groups. However, if this is not feasible (e.g., due to small intake sample in rural offices), it is acceptable to put a smaller proportion into the comparison group. With unequal group sizes, however, more study participants are needed to achieve the same level of power. Given the number of participants available in most states, sample size is unlikely to be a substantial problem. Attachment II provides a sample size calculator that can help determine the sample sizes needed for the treatment and comparison groups.

F. Summary of REA Process

Above, we presented various issues associated with designing and implementing a rigorous impact evaluation of the REA Initiative. In Figure 2, we present a summary of processes that can lead to a successful implementation of the REA Initiative and provide the necessary sample

for a rigorous impact evaluation of REA services. We recommend that states use this process if possible.

Figure 2: REA Recommended Process



For states that adopt this process, individual sites may adjust the number of treatment group members who are not invited to participate in REA. For example, if there are 50 treatment group members on a weekly list, sites that can only accommodate 30 REA interviews per week may invite the first 30 cases for an REA interview. Sites should maintain a record of which treatment group members were selected for an invitation. The remaining 20 cases from that week should be (1) removed from the pool of eligible treatment group members; (2) should not be invited for REA services, and (3) should not be included in the comparison group. The entire process should be repeated with each selection of REA participants.

G. Best Practices for Implementing the Random Assignment Design

In this section, we present some best practices that states should consider when implementing a random assignment design. These best practices have been developed from the experiences of a number of states.

- ***Keep the design simple and easy to implement***. We recommend a simple random assignment design. This simple design is generally easier to implement and maintain than more complex designs. A number of states have attempted to introduce additional complexity to this basic design. With additional complexity comes implementation difficulties.

- ***Apply the same eligibility criteria for selection into the comparison group as used for the REA treatment group***. To create a comparison group that is comparable to the REA treatment group, both groups should be drawn from the same population. Some states have created different criteria for the treatment group and the comparison group. For example, in some states, eligibility for participation in the REA Initiative is subject to certain requirements such as: no return to work date, availability for full time work, length of UI claim, absence of non-monetary issues affecting the claim, etc. If the same criteria are not applied to the comparison group, sampling bias may be generated. In summary, all eligibility criteria established for participation in the REA Initiative should *also* be applied for inclusion in the comparison group.

- ***Use random assignment procedures to assign claimants to REA treatment or the comparison group***. A critical aspect of random assignment designs is that every person in the pool of possible participants has the same chance as any other to be selected into the treatment or comparison group. Any characteristic that is related to both placement into the treatment or comparison groups *and* to the intended outcome measures can bias the results. Since many factors can be related to the outcomes, sometimes in non-intuitive ways, the best way to ensure that selection into groups is unrelated to outcomes is to use a randomization procedure.

To avoid any potential biases in the formation of the REA participant and comparison groups, states should use random assignment. One method is to use the last digit of the claimant's social security number, which is essentially random, to form the groups. Since the last digit of an individual's social security number is random, assigning eligible claimants whose social security number ends in 0-4 to the treatment group and the rest (i.e., SSN ending in 5-9) to the comparison group will generate equal sized treatment and comparison groups.

- ***Eliminate discretionary selection factors for creating the REA or comparison groups.*** Because it is critical that the factors used for selection into the treatment and comparison conditions are unrelated to the outcomes, the use of any factors left to staff member discretion can inadvertently bias the results. Well meaning staff may believe that one type of person "needs" REA more, or that another type should not be assigned. Over time, such decisions would result in biased results, e.g., claimants more difficult to reemploy might be systematically assigned to REA thus biasing the results against the treatment (or vice versa). For this reason, only rigorous random assignment procedures should be used to determine whether an eligible participant is assigned to the REA treatment group or to the comparison group.

- ***Include in the REA group anyone invited to participate***. When determining who should be counted as a member of the REA treatment group, the state and local area should include anyone who was invited (or instructed) to participate. The invitation itself is a part of the REA treatment. Therefore, anyone receiving the invitation should be counted as in the REA group, whether or not they actually participate in an REA session.

- ***Select REA and comparison group members from the same locations/offices.*** REA and comparison groups should be selected from the same locations. Many states have implemented the REA Initiative only in particular locations due to resource limitations and the need for sufficient population to meet the program goals. To keep the REA and comparison groups equivalent, it is preferable to draw the comparison group from these same One-Stop Career Centers as the treatment group.

Although it may be more convenient to draw REA participants from one set of locations, and use claimants from other locations as a comparison, this method creates biases. Often different locations serve clients with different demographics, different occupational mixes, or are differentially affected by specific events impacting the local labor market (e.g. a plant closing). To ensure the REA and comparison groups are both impacted equally by such factors, both should be drawn from the same offices.

If it is not possible to select REA and comparison group members from the same location (e.g., all claimants in one or more of the participating offices are served by REA in order to reach program participation goals) claimants from other local offices can be used, although doing so creates the potential for many biases. In this situation, establishing in advance that the offices selected for the comparison group serve similar populations, occupation mixes, and have similar outcome characteristics will be necessary to rule out potential biases based on location differences. For example, pre-REA comparisons should be performed to establish the comparability of the REA and comparison sites, such as the demographics of claimants, the occupation mix of claimants, average claim duration, average dollars received, disqualification rates, etc.

- ***Select claimants into treatment or comparison groups at the same time.*** Because time is related to relevant UI outcomes (duration, amount received, etc.), if the timing of selection into treatment or comparison groups is not the same, bias can result. For example, if the comparison group is selected immediately, but the REA group is not selected until the third week, the comparison group can contain people that obtained work in the first two weeks of the claim, but the REA group cannot. This can artificially lower the claim duration average for the comparison group. Selecting the comparison group at a different time than the REA participants are selected can introduce bias, therefore, the treatment and comparison groups should be selected at the same time.

  Some states have developed an REA design that keeps eligible claimants in a selection pool for a specified period of time. In these states, if an individual is not assigned to the

treatment group during this period, the individual is assigned to the comparison group. This design violates the requirement that treatment and comparison group members should be assigned at the same time.

A variation of the above design has been developed and used in other states. In this variation, comparison group members are randomly selected immediately after they file a claim while treatment group members are assigned into a pool for later selection. This design also violates the requirement that treatment and comparison group members should be assigned at the same time.

- ***If the state's REA initiative includes more than one REA per claimant, consider them all as part of the REA treatment.*** If the REA claimants are offered multiple REAs during their claim year, it is not necessary to select multiple treatment and comparison groups. Having one or more REAs is simply considered part of the overall "REA treatment" as implemented by the state. The design can still include one treatment group and one comparison group as per the recommended design. The comparison group should be selected at the time the REA treatment group is selected for their *first* REA. States electing to provide multiple REAs to claimants must ensure that each REA includes each of the required elements of an REA.

- ***Perform the selection into the REA treatment and comparison groups once, and at the same time.*** The simplest procedure for selecting the treatment and comparison groups is to have one selection into REA or comparison groups, done at the same point in the program. More complex designs that involve multiple opportunities for selection into the treatment or comparison group are more complex to implement, and provide more opportunity for introducing biases.

  However, if the state must provide claimants with multiple opportunities for selection into the REA group (e.g., selection into REA is done weekly, and a claimant not selected in week two may be selected in subsequent weeks to meet program participation goals), they should likewise have multiple opportunities to be selected into the comparison

group. Otherwise, the longer a claimant is unemployed, the more likely they are to be placed in the REA group, but not the comparison group, thus potentially inflating the average claim duration for the REA group relative to the comparison group due to selection bias.

If multiple opportunities exist for an individual to be selected for the REA treatment, then the same multiple opportunities must exist for that individual to be selected into the comparison group. For example, if the state's Initiative is designed such that some people receive their first REA in week two of their claim, whereas others receive their first REA at weeks three, four, or five of their claim, then the appropriate proportion of comparison group members should also be selected in weeks two, three, four, and five of their claim.

- ***Ensure a sufficient number of claimants in each group to provide sufficient power to detect the expected effect size.*** The number of claimants required for an evaluation depends on several factors: 1) the size of effect the state wishes to detect, in this case, the difference in average claim duration between the treatment and comparison groups; 2) the level of confidence the state is seeking to determine that any significant effect found is not just an anomaly; this is called the alpha level; 3) the power with which the state wishes to be able to detect a real effect if one exists; and 4) the standard deviation of the variable of interest in the population, in this case claim duration. DOL has recommended, as a goal, that the sample size over a one-year period should be at least sufficient to find a difference in average duration of claim of 0.3 weeks statistically significant at an alpha of 0.05, with power of 0.8. Generally speaking, given the number of people that are proposed for REA treatment by most states, sufficient sample size is not likely to be problematic.

- ***An individual claimant can be placed in either the REA treatment group or the comparison group, not both.*** Once a claimant has been selected for the treatment group, whether they receive any REA services or not, they should not subsequently be placed into a comparison group. That is, once a claimant has been notified that they are to

receive an REA, they have potentially been impacted by the REA process. Therefore, they cannot be used in any comparison group. Likewise, an individual should not be selected for an REA after having been previously chosen as a comparison group member. In summary: once a claimant is a treatment (comparison) group member, that claimant is always a treatment (comparison) group member.

Attachment I

Quasi-Experimental Approaches – Analytic Methods for Creating a Comparison Group

Under some extreme circumstances a state may simply have no available claimants to construct a viable comparison group or may not have allowed for the proper selection of a comparison group and wishes to retroactively construct a group. There are several statistical methods that allow for the construction of a comparison group chosen from the pool of state claimants who did not participate in the REA Initiative.

These methods are complex and difficult to implement. These quasi-experimental methods use statistical procedures to create an appropriate comparison group that is similar to the treatment group. While these procedures reduce the likelihood of sample selection bias, random assignment design (if feasible) is still the preferred approach. The complexity of properly employing these alternative methodologies requires a researcher with experience and expertise to select the right technique and to ensure that the technique is implemented properly.

One method, in particular, has frequently been used to define a viable comparison group under these circumstances: Propensity Score Matching. This method is described in many statistical and econometric textbooks and is included in a number of statistical software packages (SAS, and SPSS). A brief description of this method is presented below.

Propensity Matching Score

This method, introduced by Rosenbaum and Rubin (1983), is used most often to provide an alternative method for estimating treatment effects when treatment assignment is not random. In propensity score matching, an algorithm is used to select the best match for each treatment group member from among the available potential non-participants. Logistic regression is used to derive the probability of participating based on relevant variables. Each person in the treatment group is then matched with one or more non-participants with a similar likelihood of participation but who were not selected.

Attachment II

Power Test Calculator for the Minimally Sufficient Number of Members in the Comparison Group

When using a random assignment design to measure program impacts, the most efficient allocation of the sample is half treatment and half comparison. If this allocation is not practical (e.g., if the selected sites are too small to provide the required number of REA participants), the state may select a smaller comparison group sample. To determine the minimally sufficient comparison group size, we recommend that states use a power/sample size calculator. The power/sample size calculator below has been selected for its simplicity and the flexibility in adjusting parameters. The power/sample size calculator can be found at: http://www.stat.uiowa.edu/~rlenth/Power/index.html.

Below are directions for finding the minimum comparison group size.[3] The directions are followed by an example.

1) Go to the website above.[4]
2) Select "Two Sample t-test (pooled or Satterthwaite)" and click "run selection." The java application will open.
3) Select "solve for sample size" in the lower right.
4) Select power of .8 (to be precise in entering numbers, it is helpful to click the small shaded area to the top right of the slide bars and enter the value directly, *then click OK to set the value*)
5) Set the Alpha value at .05 (p=.05 test) and check "two-tailed"
6) Set the "true difference of means" to .3 (for the minimum difference in claim duration we wish to have the power to detect as a significant difference).
7) Then use an estimate of the standard deviation (SD) of claim duration in weeks to estimate the SD and enter that. Check the box for equal sigmas.[5]
8) Under Allocation, select "independent" from the pull down menu.
9) Set the "n1" to the sample size of the treatment group using the slide bar or value entry

[3] Lenth, R. V. (2006). Java Applets for Power and Sample Size [Computer software]. Retrieved July 10, 2008, from http://www.stat.uiowa.edu/~rlenth/Power.

[4] Please note that depending on your computer's software, it may be necessary for you to download a more recent version of JRE and/or use an older version of the calculator as indicated on the website.

[5] You can generally get this estimate based on pre-REA data. If you have an estimate of the expected SDs for each group, you can select "unequal" and set them independently. In this case, if you want to know the optimal sample size for each group given these SDs, you can set Allocation under the "n2" box to "optimal" instead of independent and it will provide the sample sizes adjusted for expected SD of each group.

10) Set “n2” to “minimum” using the pull down menu

11) You may have to enter and adjust sample size for n1 and power a few times. Be sure to click on “ok” after entering or changing a value. Once power is at .8 and n1 is the size of the first group, n2 will provide you with the minimum required comparison group.

Example

For this example we assume an REA treatment group of 10,000 as determined by the program goals. We want a power no less than .8 to detect a minimum difference between the treatment and comparison groups means of .3 weeks of duration. The alpha level will be the standard p=.05 (two tailed test). For this example, we will say that we have found that the standard deviation in duration has been 5 weeks, and we assume this to be true for both groups. When we set these parameters, we find that the minimum comparison group needed is 2,789.

Click here to change from slide-bar to direct entry.

This will be the sample size needed for the comparison group.