Vocational Rehabilitation Program Evaluation Coach

TOOLKIT

Table of Contents

VR Program Evaluation Coach

Vocational Rehabilitation Program Evaluation Overview

INTRODUCTION

In an ideal world, as you decide whether to implement or continue offering a vocational rehabilitation (VR) program, policy, or service, you would base your decision on how well the service meets your needs and improves outcomes for your clients. Mathematica Policy Research and the U.S. Department of Education's Rehabilitation Services Administration are supporting program evaluations, including quick-turnaround rapid cycle evaluations, to help VR agencies test and evaluate services and service changes. The findings generated through the rapid cycle evaluation process can help you make decisions by providing you with evidence on how well a service works.

We have built a free, web-based toolkit—the VR Program Evaluation Coach—that will guide you through the steps of conducting a program evaluation. As you move through the VR Coach, the tools will help you to design and conduct your own program evaluations and pilot tests.

The VR Coach can help you answer important questions such as:

- Does a new program lead to outcomes that my agency and our clients want?
- Should we keep paying for a service we're already using?
- How can we set up meaningful pilots of service changes?

Mathematica Policy Research prepared the VR Program Evaluation Coach under contract to the Institute for Community Inclusion (ICI) at the University of Massachusetts-Boston as part of grant number H235LI00004 awarded by the U.S. Department of Education, Rehabilitation Services Administration.

Guide to Using the VR Program Evaluation Coach

VR Program Evaluation Coach

What You'll Need to Get Started

This document outlines some basic requirements for conducting your own program evaluation and the steps you will take with the Vocational Rehabilitation (VR) Program Evaluation Coach. Along the way, you might have to bring in other members of your team (such as a data analyst) to help.

Before you can begin crafting a program evaluation with the VR Program Evaluation Coach, you need the following:

- A service to test (may or may not already be in use)
- A population receiving or willing to receive the service
  - Within the population of recipients or potential recipients, you will have to create two similar groups, one receiving the service and one not receiving the service. The VR Program Evaluation Coach will help you create the two groups.
- A program for collecting and organizing data, such as Microsoft Excel

REQUIRED INFORMATION

Exhibit 1 summarizes the information you will need to complete various steps for the program evaluation. The earlier you begin collecting and thinking about the information, the easier it will be to complete your program evaluation.

Exhibit 1. Information needed for the program evaluation

Step	Information needed
The basics	• Who the service recipients are (or will be) • What outcomes you're interested in
Determine your program evaluation approach	• Whether the service has been implemented • How you will sort potential recipients into similar groups
Craft your program evaluation question	• What outcome you are targeting • Whom you are trying to affect
Making sense of your results	• Cost of the service • Idea of what success looks like
Summarize context	• Basic information about the service and its implementation • Details about the VR environment

VR Program Evaluation Coach

REQUIRED DATA

The earlier you begin collecting and cleaning your data, the better. A review of our data guides, "Prepare Your Data for Analysis: Matched Comparison Design", (https://vrevalcoach.com/static/pdf/Prepare_Your_Data_for_Analysis_Matched_Comparison_Design.pdf) and "Prepare Your Data for Analysis: Randomized Pilot", (https://vrevalcoach.com/static/pdf/Prepare_Your_Data_for_Analysis_Randomized_Pilot.pdf) will help acquaint you with the format required for the data you will upload to the VR Program Evaluation Coach and how to ensure that you enter data in that format. Depending on the type of program evaluation you are conducting, you will have to use data at different points. It is likely that you won't be able to gather all of your information at once, but the earlier you complete your data set, the easier your analysis will be. Note that it is important for you to record all of your data using individual identifiers. These are unique codes for each participant. The identifiers—individual, staff, or office identifiers—are what will enable you to combine (merge) data sets without using any personally identifiable information (PII). The VR Program Evaluation Coach does not require personally identifiable information for use of the tool, and it does not create evaluation reports or briefs that contain PII. Below is a list of the data you will need (for more details, refer to the data guide linked above).

- Outcome data (required)
- Outcome or outcome-related data from before use of the service (optional)
- List of who is and is not receiving the service (required)
- Background characteristics (important for establishing a comparison group)
- Service delivery data (required for certain program evaluation questions)

If you are not directly managing the data for the program evaluation, the planning stage is an important time to communicate with the person responsible for data management. You should discuss the data required for your specific program evaluation and determine how and when those data will be collected.

REQUIRED KNOWLEDGE AND SKILLS

Throughout the course of your program evaluation, you and/or someone on your team will have to possess the following skills and competencies:

- Ability to answer questions about the service being evaluated and the desired outcomes
- Understand how to collect or extract data from different sources
- Proficiency with a data management system

VR Program Evaluation Coach

ICON DIRECTORY

Guide: These documents are meant to assist you in completing your program evaluation by providing you with educational material about the program evaluation process.

Interactive tools: These tools require you to input information relevant to your program evaluation.

Statistical tools: These tools analyze the data you input to determine who is in your treatment and comparison groups and whether your service is achieving the desired outcome.

VR Program Evaluation Coach

Choose a Service

Innovations in vocational rehabilitation (VR) programs, policies, and services can help to achieve the goals you've set for your VR offices or counselors. But selecting the service that best fits your needs can be challenging. The VR Program Evaluation Coach is targeted to VR agencies that have already selected or implemented a service change. If you are looking for a service to implement, this guide will help you with that process, and you can return to the VR Program Evaluation Coach when you are ready to plan a pilot of that service and start your program evaluation.

THREE STEPS FOR SELECTING THE RIGHT SERVICE

1. Conduct a needs assessment.
2. Discover what service might fit your needs.
3. Vet potential services.

1. CONDUCT A NEEDS ASSESSMENT

A needs assessment will help you identify gaps (needs), set priorities, and determine important criteria for solutions—the services you choose. The assessment can take many forms. It can be informal conversations or interviews with various staff, counselors, or administrators, or it can be a more formal survey, focus group, or working group. You will have to decide what type of needs assessment you would like to undertake based on the time and resources you have available. If you feel as though you already know the needs of your agency and have a sense for how service changes can help you meet those needs, you might not have to conduct a needs assessment. But if you choose to conduct a needs assessment, we've included some helpful guidelines:

A. **Determine your knowledge objective(s).** It's important to outline clearly what you want to learn from the assessment you are conducting. Write out some questions that you want to be able to answer at the end of the assessment based on the information you collect. This will help you determine whom to talk to and what questions to ask them. For example, if you want to know what features might be most effective with a youth transition program, you might want to talk to students, school staff, and VR staff about their needs and about what they think is missing in existing services. If you want to know if you should target specific subgroups of high school students with the service, you might want to talk to VR staff about their needs and about which students need the most help. Make sure to write out each of these knowledge objectives and refer back to them throughout the process of designing and conducting your assessment.

B. **Choose whom you want to participate in your assessment.** Before you begin conducting your assessment, you should determine whom you want to hear from. Perhaps your objectives require input from a wide variety of sources, or maybe you want to focus only on VR staff, administrators,

or clients. Determine whom you'll need to speak to and invite them to participate in your assessment.

C. **Develop your questions.** When you identify your participants, develop the questions that you will ask them to reach your knowledge objectives.

D. **Conduct your assessment.** Now you can begin your assessment. You'll want to record the responses. For an informal assessment, this could mean taking notes during or after your conversations. For a more formal survey, this might mean recording responses in a systematic way.

E. **Analyze your data and draw conclusions.** After you've spoken to everyone you identified, look back at your data (notes or survey responses) for trends and themes. Have a number of participants made similar comments? Are there needs that people already agree on? These are the items you'll want to consider as you start looking for the right service change.

2. DISCOVER WHAT SERVICE MIGHT FIT YOUR NEEDS

Here are some questions to keep in mind while you consider various services:

- What problem are you trying to solve, or what opportunity do you wish to pursue?
- Who should be involved in the selection process?
- What are the technical requirements?
- How much training is required, and how will it be delivered?
- What's the intended implementation approach?
- What type of data or reporting tools do you need?

3. VET POTENTIAL SERVICES

In addition to existing requirements and discoveries from the needs assessment, here is a checklist of questions to consider when vetting services:

- ☐ Does the implementation model you have in mind match what the service is designed to do?
  - Counselor-level versus office-level services
  - Students or youth versus working-age clients
  - Counselor-provided services versus purchased services
- ☐ Service fit
  - What problem do you wish to solve, and how does the service address the problem?
  - Does the service address the desired employment or other outcome(s)?
  - Does the theory of change or learning underlying the service match your approach?
- ☐ Implementation issues
  - How easy is it to implement or use the service?
  - Do VR counselors and staff have to participate in some preparatory training?

- Is the service compatible with your agency's existing policies and programs?
- For services meant to be used outside the office, do intended users have access to the needed resources?

☐ Cost and time needed for implementation

- What costs are associated with implementing the service? Are any additional equipment purchases involved?
- What costs are associated with administrative and staff time for training, learning about the service, and incorporating the service into the existing milieu?

☐ Evidence of effectiveness

- Is there strong evidence of the effectiveness of the service?
- Is the evidence of effectiveness specifically for clients and settings like yours?

☐ System data availability

- What learning and usage indicators does the service require?
- Do you have the information you need to determine whether the service is effective in achieving your goals?

VR Program Evaluation Coach

The Basics

The Basics enables you to enter simple information about your service change. It is the information that the Vocational Rehabilitation (VR) Program Evaluation Coach will use throughout your program evaluation to personalize and refine your experience. Here is a guide to the foundational questions included in The Basics; you can use this document for additional guidance on how to answer these questions.

Question	Explanation	Example
Have you decided on a service change that you would like to test?	The VR Program Evaluation Coach is designed to help you evaluate a service or policy change. To begin a program evaluation and answer all subsequent questions in the VR Program Evaluation Coach, you have to know what you are evaluating. The VR Program Evaluation Coach is not designed to help you select or compare potential services to be evaluated, but you can refer to our brief guide, "Choose a Service", (https://vrevalcoach.com/static/pdf/Choose_a_Service.pdf) if you have not yet selected a service to evaluate.	You may select only yes or no for this question.
What is the name of the service or program you are testing?	This refers to the name of your service or program. By answering this question, the VR Program Evaluation Coach will be able to ask about your specific service or program as you move through your program evaluation. The name will also be used in your findings brief.	Example: Motivational interviewing training
What is the outcome that you want the service change to achieve?	This is a broad question about your goals for the service change. It helps the VR Program Evaluation Coach modify future questions to reflect the outcome of interest. In a later step, you will provide more detail on how you will measure the outcome and the specific changes you want to see. You can find examples below for how to categorize your goals.	Example: Employment
Who is or will be the target group for the service change?	This question refers to the people directly engaging with the service or program. In some cases, several groups, such as both clients and VR staff, will be recipients of the service. You should select the recipients for whom you are interested in measuring outcomes. On the other hand, if you think it is important to identify all recipients, you can select other and type in "clients and VR staff" or another combination of recipients.	Example: Clients

VR Program Evaluation Coach

HOW TO SELECT AN OUTCOME

Outcomes are knowledge, skills, attitudes, or other desired benefits attained as a result of an activity (for example, implementing a change in VR policies, programs, or services). We encourage you to meet with your team and determine the outcome you are most interested in measuring during your program evaluation. When you have selected an outcome, you will have to categorize it in one of three broad categories. Some of the examples below can help you choose which category to select. If none of the categories covers your outcome, you can select other and enter a description.

Employment outcomes	Choose this option if you are using a service to improve the ability of clients to get a job, the types of jobs that clients secure, the amount of clients' hourly wages or weekly earnings, or the rate of clients exiting successfully from VR.
Service outcomes	Choose this option if you are using a service to help more clients obtain a signed IPE, obtain a signed IPE more quickly, receive specific services (such as pre-employment transition services or job placement services), speed up the timing of client service receipt, or improve referrals to providers.
Skills outcomes	Choose this option if you are using a service to improve clients' skill development or measurable skill gain or to help clients enter or complete a training program or obtain a high school diploma or postsecondary degree.
Other outcomes	Choose this option if you are using a service to improve other outcomes that do not fall under the three broad categories, such as effectiveness in employer outreach, secondary school referrals, or counselor training.

VR Program Evaluation Coach

Determine Your Program Evaluation Approach

There are several ways to conduct a program evaluation for your vocational rehabilitation (VR) agency. Before you can begin working through your program evaluation, we'll help you find the approach that works best for you. The decision tree provided here will help you understand the pathway that suits your unique context. Please note that, by answering the questions in this step of the VR Program Evaluation Coach, you can rely on the VR Coach to determine the best evaluation approach for you.

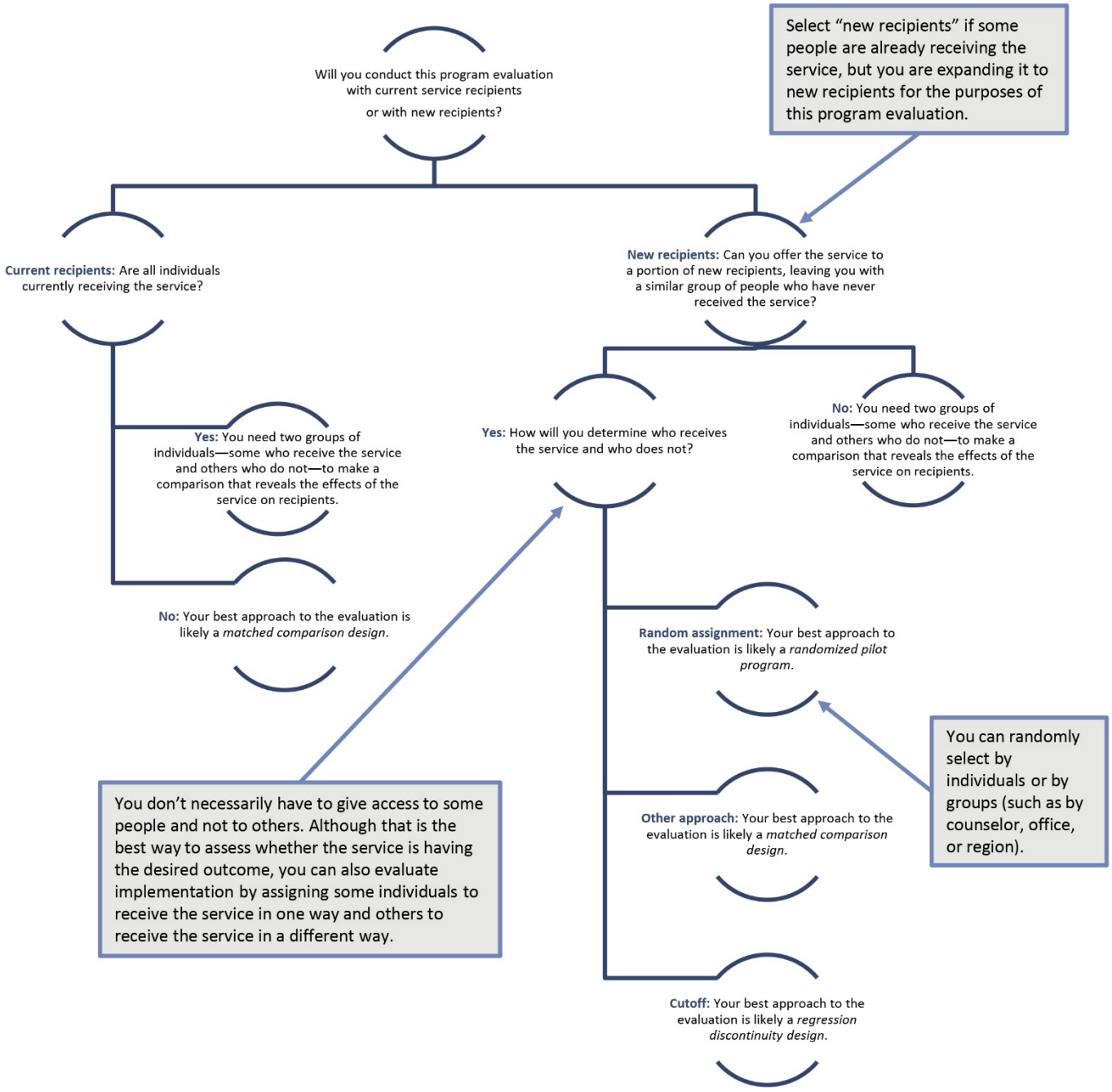
The VR Program Evaluation Coach is a powerful tool that helps you move quickly to make decisions based on rigorous evidence—but you'll have to gather some data to get there. It is never too early to start thinking about how you will conduct your evaluation, how the approach you choose will affect your data collection needs and efforts, and whether you'll have to bring in any additional team members to help with the data collection and evaluation process.

Note that there are three potential approaches to program evaluation (two of which are currently supported by VR Program Evaluation Coach):

1. Randomized pilot
2. Other comparison group/matched comparison design
3. Cutoff/regression discontinuity design (not supported by VR Program Evaluation Coach)

Please follow the decision tree in Exhibit 1 to determine the approach that makes best sense for your needs and context. The decision tree will also help you understand what changes you can make to ensure that you can continue with either a randomized pilot or other comparison group/matched comparison design.

Exhibit 1. Decision tree



CHOOSING AN EVALUATION METHOD

The method you choose will affect how confident you can be that your service leads to the outcomes you observe and that your findings will apply to a broader population.

Random assignment—most rigorous

Random assignment is similar to putting the names of individuals in a hat and randomly picking some of the names. If you put in 10 names and select 5, those you pick would get to pilot (or try) the service change, and those not selected would continue to receive your existing services. In this scenario, each individual would have a 50 percent chance of selection. The VR Program Evaluation Coach can conduct the process for you by using a digital hat and show you the similarities between the groups. The process can work with clients, VR staff such as counselors, offices, or providers.

Advantages of random assignment

- The resulting treatment and comparison groups will, on average, be similar in both observed (for example, gender) and unobserved (for example, motivation) characteristics, giving you the most confidence that any differences in outcomes between the groups are attributable to their receipt of the service.
- Given that everyone has the same chance of selection, random assignment is a fair way to decide who gets to receive a service.
- Findings are relevant to the overall population with which you are working (that is, everyone in the hat).

Disadvantages of random assignment

- The random assignment evaluation approach is the hardest to implement.
- Some individuals get to receive the service earlier than others, which might be a problem for some stakeholders and might be contrary to the need to offer services statewide.

Other comparison group/matched comparison design

With the other comparison group/matched comparison design, assignment to the treatment and comparison groups takes place in a nonrandom manner based on specific criteria. For example, the offices assigned to the pilot or treatment group may be those in a specific location, or VR counselors assigned to the pilot group may be those who will deliver the service because of their access to providers, or current clients may be compared to similar clients who received services before the introduction of the service change.

Advantages of other comparison group/matched comparison design

- The design might align well with how you've assigned service receipt in the past.
- The approach might be logistically easier, particularly if you're selecting participants based on who has easier access to services or populations.

Disadvantage of other comparison group/matched comparison design

- Stakeholders might not consider the service assignment process to be fair. The individuals that meet the criteria might already have advantages over other individuals or groups.
- The process might not align with need. Those who need the service more or who would gain more from it might be less likely to receive it than if you used random assignment.
- It's not always possible to create a good comparison group. If the VR Program Evaluation Coach fails to create similar groups of service recipients and nonrecipients, you will not be able to determine whether the service is having the desired outcome.
- It's also possible that groups appear similar but that unobserved differences not included in your data are what drive the outcome—not the service change.

Cutoff/regression discontinuity design

Using a cutoff means selecting who receives the service based on where individuals rank in relation to an existing threshold. For example, students just below a "low score" threshold are assigned to receive the service, or clients who apply for a service just before or after the date that order of selection takes effect. We assume that individuals close to a cutoff are very similar; therefore, you can compare those just below the cutoff with those just above it.

Advantage of cutoff/regression discontinuity design

- A cutoff enables you to compare similar recipients without using random assignment.

Disadvantages of cutoff/regression discontinuity design

- Findings apply only to individuals near the cutoff, not to the whole population.
- The process might be less fair than random assignment. With this method, scoring or ranking one point below the cutoff gives you a 100 percent chance of receiving the service, whereas scoring or ranking one point above the cutoff means you have a 0 percent chance of receiving the service, even though both groups of individuals might be equally in need of it.
- The VR Program Evaluation Coach does not currently support the cutoff/regression discontinuity design. Contact us if you're interested in testing this method.

NOTE: Across all of these methods, you do not have to create or compare groups at the individual level. If there are logistical barriers to working with individuals, you can create groups at higher levels, such as groups of counselors, offices, regions. However, at higher levels, you will need more study participants to be confident in your results.

VR Program Evaluation Coach

Craft Your Program Evaluation Question

A carefully crafted program evaluation question grounds the development of a program evaluation in a clear statement of the evaluation's goals. This guide highlights the key components of a well-defined program evaluation question and provides examples for each of the four types of program evaluation questions that you can answer by using the Vocational Rehabilitation (VR) Program Evaluation Coach.

INGREDIENTS OF AN EFFECTIVE PROGRAM EVALUATION QUESTION

Exhibit 1 illustrates the key ingredients of a focused program evaluation question, or the "**A, B, C**, and **D**s." The blueprint for an effective program evaluation question is structured as follows:

Does **A** lead to **B** among **C,** compared with **D**?

Careful consideration of the details of the intervention; its intended goal; and the target population of clients, VR staff, or providers will make the rest of the evaluation process much easier. However, you might not yet know every aspect detailed in Exhibit 1. If that's the case, the exhibit can help identify critical aspects of the program evaluation that you will want to discuss with key stakeholders as soon as possible.

EXPECTED ANSWERS TO A WELL-FORMED PROGRAM EVALUATION QUESTION

A program evaluation provides one of three answers:

1. Yes, **A** is likely to lead to **B** among **C**, compared with **D**.
2. No, it is not likely that **A** will lead to **B** among **C**, compared with **D**.
3. More data are needed to reach a strong conclusion.

The answer could drive important decisions within your agency. The VR Program Evaluation Coach's guide "Making Sense of Your Results", (https://vrevalcoach.com/static/pdf/Making_Sense_of_Your_Results.pdf) will help you think through how you will use the evidence you uncover through the evaluation process. Given that a well-designed program evaluation aims to provide a clear answer to the precise question it was designed to address, it is important to confirm with program evaluation stakeholders that the question addresses a learning objective that can help direct decisions and actions.

Exhibit 1. Components of a well-crafted program evaluation question

The blueprint: Does A lead to B among C, compared with D?	
"**A**" is the name of the service you are testing. Consider including how often and for how long recipients interact with or are exposed to the service.	**Examples** • Motivational interviewing training • Enhanced job placement services • Financial planning tool
"**B**" is the intended effect of using the service. To identify "B," you will have to determine your outcome of interest (what you hope to change by using the service), the direction of the intended change (increase or decrease), and the measure you will use to determine the effect (if applicable). An outcome is knowledge, skills, attitudes, or other desired benefits that are attained as a result of an activity (for example, implementing a change in VR policies, programs, or services). We encourage you to meet with your team and determine the outcome you are most interested in measuring during your program evaluation.	**Examples** • Increased rate of employment for youth clients after receiving services • Decreased average time between application and signed Individualized Plan for Employment (IPE) • Increased number of high school students receiving pre-employment transition services • Increased number of referrals to providers for a specific service
"**C**" is the group of people for whom you want to see results. What group of people are you trying to affect? For example, clients might be receiving the service to increase their rate of exiting with employment, or counselors might be using it to increase the number of eligible applicants who sign an IPE. In both cases, the answer to Part C is the clients.	**Examples** • Transition-age youth not in school • Individuals with specific disabilities (such as mental illness, learning disabilities, or autism) • Adults who are working at the time they apply for VR services
"**D**" is the group of people against whom you want to compare the group defined in "C." This group should be as similar as possible to those identified in "C." In an ideal world, the only difference between the two groups would be receipt of the service. When thinking about "C," consider where the service is used. If the service is used by staff in three VR offices, then clients or staff in other VR offices might be an appropriate comparison group.	**Examples** • Similar adults with autism whose counselors did not receive additional training in motivational interviewing • Similar pre-employment transition services students who receive only workplace readiness training instead of workplace readiness training and job exploration counseling

Program evaluation question examples

The VR Program Evaluation Coach is designed to help answer four key types of program evaluation questions. Exhibit 2 lists examples of each type.

Exhibit 2. Examples of program evaluation questions, by type

Category of program evaluation question	Examples of clear, narrow program evaluation questions
Does this service achieve its intended outcomes?	• Does receipt of work-based learning experiences by VR clients enrolled in high school increase employment outcomes at exit compared with VR clients enrolled in high school who do not receive those experiences? • Does receipt of work-based learning experiences by VR clients enrolled in high school increase high school degree attainment compared with VR clients enrolled in high school who do not receive those experiences? • Does receipt of work-based learning experiences by VR clients enrolled in high school increase employment outcomes at exit compared with VR clients enrolled in high school who receive services through another program targeted to VR clients enrolled in high school?
Does this training help recipients engage with this service more and/or better?	• Does providing high school staff with training for pre-employment transition services improve the proportion of high school students receiving those services compared with high school staff who did not receive the training?
Does providing information to clients or VR staff change behavior?	• Does exposure to a dashboard summarizing client services and outcomes result in more VR counselor outreach to disengaged clients compared with VR counselors who did not have access to a dashboard? • Does VR counselor receipt of weekly reports on client times to IPE increase the number of clients who obtained IPEs compared with clients whose VR counselors did not receive weekly reports?
Does this modification to the service (or how it is implemented) make it work better?	• Do work-based learning experiences provided to VR clients enrolled in high school by VR staff result in higher completion rates of those experiences compared with work-based learning experiences provided to VR clients enrolled in high school by community rehabilitation providers? • Do work-based learning experiences provided to VR clients enrolled in high school during the summer result in higher completion rates of those experiences compared with work-based learning experiences provided to VR clients enrolled in high school during the school year? • Does refresher training to VR counselors in motivational interviewing lead to higher rates of IPE completion compared with VR counselors who did not receive refresher training in motivational interviewing?

Making Sense of Your Results

We'll use this tool to think about how the answer to your program evaluation question will guide the decisions you make about your service change. The following questions will help you prepare to think about the relationship between your results and the service costs and stakes associated with your decision. To complete this tool, you will have to consider the following:

What would success look like?

How much risk are you willing to take when making a decision or a recommendation?

CONSIDER: When making decisions, it is important to think about the costs of the service you're considering and the risks associated with your decision.
For example, you might try to decide:

- Whether to continue using a provider you already use
- Whether to purchase a specific type of training for all vocational rehabilitation (VR) counselors
- Whether to change how VR counselors serve their clients, based on a service tested with your program evaluation

A. WHAT IS THE COST OF THE SERVICE?

The cost of a service will probably be a factor in your decision about how great its effect needs to be for it to be a good value. Cost might also influence how certain you want to be that the service has the desired outcome.

The cost of a service could simply be the cost per recipient. It could also be the savings you realize by replacing one service with another less expensive service. Cost can also be measured in something other than dollars, such as counselor time saved.

How much do you pay (or save) per recipient by using the service?

Examples

- It costs about $20 per pre-employment transition service student to provide instruction in self-advocacy.
- It saves about $1,000 per office to have a job developer conduct a specific assessment rather than relying directly on VR counselors to conduct an assessment.
- We're not sure. We believe that relying on a single VR counselor to work with high schools will decrease the time that all counselors must devote to working with schools to identify potential students for services.

B. WHAT WOULD SUCCESS LOOK LIKE?

The VR Program Evaluation Coach will calculate the likelihood that the service has an effect at least as great as the number you specify. The VR Program Evaluation Coach uses the units of whatever you are using to measure outcomes. For example, if you are looking at rates of exiting from VR with employment, the units would be percentage points. If you are looking at how quickly clients sign an IPE, the units would be days.

Consider: The number could be 0 or even a negative number. A 0 means that you would consider any increase in the outcome a success, as long as it is positive. If the service saves you a lot of money, you might be willing to conclude that any change greater than a small negative number is a success.

The VR Program Evaluation Coach will tell you, "There is an X percent probability that the service led to an increase in the outcome by Y units or more." In this step, you select the value for Y.

What is the direction of intended change? (Do you hope to increase or decrease the outcome of interest?)

What measure will you use to test whether the service produced an effect?

What unit of measurement does your outcome of interest use to measure services?

Examples

- Closure type
- Average time between application and IPE

What is the unit used to measure your outcome of interest?

Examples

- Dollars
- Days
- Percentage points

By how many units would your outcome have to increase or decrease for you to consider the service a success?

To answer this question, think about the impact size that would be meaningful in your context. If you hope to see an increase in employment rates, would a 5 percentage point increase, on average, be a meaningful increase? What about 10 percentage points? When determining a meaningful change, you should consider factors such as the maximum possible amount and current performance. A 10 percentage point increase on a measure with a maximum possible value of 100 is dramatically different from a $10 increase in weekly earnings. Similarly, if clients have a rate of 80 percent on

average, there is less room for improvement than if the average rate is 50 percent, which might affect the size of the improvement that you would consider meaningful.

C. HOW CONFIDENT DO YOU WANT TO BE?

Rarely do we have enough evidence to be nearly certain (close to 100 percent certainty) that a service produces the desired results. Most of the time, we have to make choices with incomplete evidence. Thinking ahead to your results, what likelihood would you need in order to be comfortable in concluding that the service is having the desired outcome?

As you think about the probability threshold that you would need to conclude that the service led to the intended outcome, you need to consider the stakes involved. You might want a higher level of certainty if your results will influence a high-stakes decision than if your results will influence a lower-stakes decision. For example:

- High stakes—if you are deciding whether to implement a statewide service change
- Low stakes—if you are deciding to request providers to offer a specific service

Though it might be tempting to set a high-likelihood threshold for any type of program evaluation, it is important to note that the higher the threshold, the harder it is to meet that threshold. The reason might not necessarily be that the service isn't working; instead, it could be driven by either the number of recipients you are testing the service with or your unit of measurement. With a particularly small number of recipients, you will be unlikely to achieve high confidence even if the service is highly effective.

Example

The VR Program Evaluation Coach will tell you, "There is an X percent probability that the service led to an increase in the outcome by Y units or more."

That is, there is a 91.1 percent probability that the service led to an increase in the outcome by 5 percentage points or more.

In this step, you are selecting the value to which you will compare X. If X is less than your chosen value, you would not be confident that the service is having the desired outcome. If X is greater than your chosen value, you would be confident that the service is having the desired outcome.

That is, if you selected 95 percent in this step, you would conclude that the service is not having the desired outcome because 91.1 percent is less than 95 percent.

D. WHAT WILL YOU DO IF…

The VR Program Evaluation Coach will provide you with one of the following three answers to your program evaluation question. It is important that you think about what you will do under each scenario.

1. . . .it is likely that the service has the desired outcome. . .
2. . . .it is not likely that the service has the desired outcome. . .
3. . . .the results are inconclusive. . .

Examples

If it is likely that the service has the desired outcome. . .

- *We will introduce a new program for all youth clients.*

If it is not likely that the service has the desired outcome. . .

- *We will look for another program for youth clients.*

If the results are inconclusive. . .

- *We will collect more data to achieve a higher level of certainty about our results.*

Summarize Context

When you share the results from your study, the information you provide through the Vocational Rehabilitation (VR) Program Evaluation Coach will help your colleagues make sense of service implementation and its results. Study context is also important for understanding how the program evaluation applies to other environments.

By describing the context based on the questions below, your report will include important contextual features about the specific service used in your study. This will also make it easier to search for other reports on similar services or in similar settings.

WHAT IS (WAS) THE SERVICE?

The questions in Exhibit 1 provide additional information about the basic elements of the service you are (were) using.

Exhibit 1. Characteristics of your service

Questions	Explanation	Example
What is (was) the program evaluation period?	This is the date that recipients start(ed) receiving the service, and the date on which you'll measure the outcome(s) of interest.	Counselors began implementing Expedited Individualized Plan for Employment (IPE) on January 1, 2017, and outcomes are measured on June 1, 2017.
What type of program is (was) the service?	This is the area of need that the service is meant to affect or address.	Expedited IPE is a policy change.
What is (was) the purpose of the service?	This is a general description of the service you are implementing and its intended effect.	Expedited IPE helps counselors meet with clients early in the application process and then move them quickly toward service receipt.
What are (were) the key components of the service?	These are the different features of the service you implement, such as key aspects of delivery and application of the service.	Expedited IPE involves VR counselors and other key staff who meet with clients in person within 30 days of application to draft an IPE.
What guidelines, if any, are (were) in place for how to use the service?	This refers to any directions for implementing the service, such as best practices or advice. It could refer to how often the service should be used, when it should be used, who should use it, or other details meant to assist you with implementing the service.	The IPE meeting should be held within 30 days of application, with IPE updates conducted as needed by the client.

HOW IS (WAS) THE SERVICE DELIVERED?

The questions in Exhibit 2 provide additional information about how the service is (was) delivered and what it is (was) used for. Your answers should reflect how you plan(ned) to implement the service. Later in your program evaluation, you will have an opportunity to document and explain any implementation issues or deviations from your plan.

Exhibit 2. Characteristics of service implementation

Questions	Explanation	Example
How often and for how long will (did) recipients receive the service?	This refers to how recipients are meant to engage with the service, in terms of frequency and duration of use, throughout the study.	IPE meeting held within 30 days of application.
How is (was) the service delivered for your program evaluation?	This refers to whether individuals received the service one-on-one, in small or large groups, in a whole office, or in a whole region.	IPE meeting held individually with clients.
What is (was) the location or setting of the service?	This refers to where the service was delivered. Options include agency offices, schools, and work sites.	IPE meeting held at agency office.

WHAT IS (WAS) THE AGENCY AND SERVICE SETTING?

Different services work well in some settings and not as well in others. The VR Program Evaluation Coach provides some guiding questions (Exhibit 3) to help you provide relevant information, along with space to provide any other contextual information that you believe is important. You might not have answers to all of the questions. Answer them as best you can with the information you have. The information will help you think about how your findings apply to a wider population within your office, region, or agency.

Exhibit 3. Characteristics of the agency and service setting

Questions	Explanation	Example
Tell us about your agency and the service	Select the type of agency that is involved in the service, input the number of recipients receiving the service, choose the appropriate geographic setting, and select the state where your agency is located.	Combined/200/urban/California
Tell us about the target group for your service	Indicate if the service is targeted to clients with specific disabilities or to specific client populations.	Individuals with autism or Supplemental Security Income (SSI) recipients
Tell us about any other contextual considerations	Include features of the setting that you think might influence how well you would expect the service to work. Features might pertain to available resources (such as types of providers in the area), available staff, or the presence of other VR programs.	Counselors had difficulty applying what they learned in their training to their specific client populations

Guide to Using Randomized Pilot Approach

Random Assignment Overview

If you select pilot participants randomly (by chance), you will create two groups that are similar on observed and unobserved characteristics. When random assignment is well implemented, you may be confident that any differences in outcomes are attributable to the service you are testing. Consequently, random assignment is considered the best choice for evaluations of effectiveness and should be used whenever possible.

THE EVALUATION CHALLENGE

You want to test whether a service change is effective, but it is impossible to observe simultaneously what happens when an individual receives or doesn't receive that service. If you introduce a service and watch what happens, you could notice improvements in, for example, rates of exiting with employment. However, you *cannot* assume that the service *caused* improved employment rates. Many other factors (including previous client work experience, other client characteristics, other vocational rehabilitation [VR] programs, external factors such as an improving economy, and so on) could have contributed to the increases.

To overcome these challenges, it is important to compare a group of service recipients to a group of nonrecipients, based on the assumption that the only real difference between the groups is whether they are receiving the service. However, the process of comparing service recipients and nonrecipients brings an additional set of challenges. When we make comparisons without trying to ensure similarities between groups, it is possible that those who receive the service differ in any number of ways from those who do not receive the service. For example, clients with previous work experience might be more motivated to receive a new job placement service, but they might also be more likely to exit with employment by virtue of that experience, leading to confusion in disentangling the effect of the service from the effect of the employment experience. Either the effect of the service or the effect of the previous experience could cause service recipients to outperform nonrecipients. Those differences can make an ineffective service look effective, or vice versa.

RANDOMIZED PILOT

Solution

The best test of whether your service works as intended is to randomly select some of the potential recipients to pilot the service and others to continue receiving VR services as usual. When the groups are assigned by chance, you may be confident that you are comparing apples to apples—that the two groups are the same in every way except for receipt of the service. Then, if you see differences in outcomes (such as employment or skills gain), you may be assured that the new service is having the desired outcome.

How it works

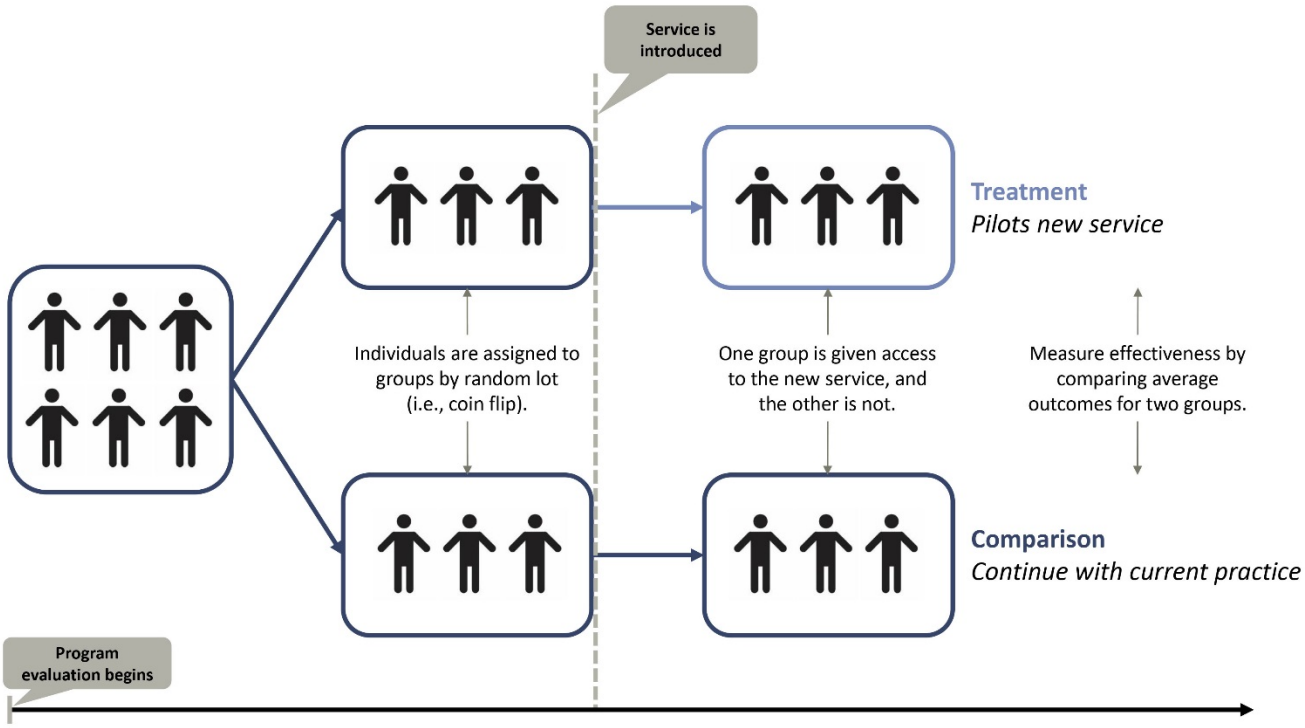
Suppose you want to determine the impact of a work readiness service on employment. To determine if the service is having an effect on employment, you would like to be able to observe your service recipients in a parallel universe. The parallel universe comprises nonrecipients who are exactly the same as recipients, but without access to the job readiness service. If the parallel group had lower rates of employment, you could conclude that the work readiness service improved outcomes for recipients.

With random assignment

Using a method equivalent to a coin toss, you may randomly assign individuals (or counselors, providers, or offices) at a point in time to one of two groups: service recipients or nonrecipients. You have two groups that, on average, are the same on observed and unobserved characteristics. Although the recipients aren't really parallel recipients, the only difference between the groups in the aggregate is whether they receive the service.

Therefore, you may conclude that any differences in achievement are attributable to the service, not to other factors.

Exhibit 1. Random assignment



BASELINE CHARACTERISTICS

An important question to ask is, Was the random assignment successful? If random assignment was successful, the two groups will appear similar on measured background characteristics, such as gender, age, and disability. A common way to assess similarity between groups is to compare the average values of the groups' background characteristics. We can quantify the difference between the two groups (recipients and nonrecipients) by using a measure called an **effect size**. Researchers use an effect size to measure different characteristics by using the same yardstick. It is calculated by dividing the difference in means between the two groups by the standard deviation of the entire sample. The VR Program Evaluation Coach's random assignment dashboard automatically calculates the effect size for any variables you specify, and it runs and reruns its randomization until you have baseline equivalence on the characteristics you selected.

The final analysis also uses background characteristics, after you implement your service and collect outcome data. It is necessary to include background characteristics in your analysis of the results if a lot of participants drop out of the study. The dropout rate in an evaluation is called attrition. For instance, if all individuals remain in the service recipient group but 25 percent drop out of the comparison group, the two groups might no longer be similar. Attrition may occur for many reasons; for example, individuals might quit a service or drop out of VR services entirely. Accounting for background characteristics can rebalance the two groups, if the initial differences were not too large. However, it is a good idea to include some background characteristics in your analysis even if your evaluation went as planned. (To learn more about attrition, see the U.S. Department of Education's What Works Clearinghouse overview of attrition standards, [http://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_brief_attrition_080715.pdf]).

Prepare Your Data for Randomization

REQUIRED

At minimum, you must include a list of participants with a unique individual identifier (ID) for each individual or group to be randomly assigned (Exhibit 1).

Exhibit 1. Unique individual identifiers

Option 1: If you are assigning individual clients, counselors, providers, or offices, your data set should include an individual ID for each.

CustomerID
159508
694677
807588
482489
555123
124226
232721
834305
490514
573401
275321
288475

Option 2a: If you are assigning groups of individuals (such as clients under counselors or counselors under offices), your data set should include individual IDs as well as group IDs.

CustomerID	OfficeID
159508	100
694677	100
807588	100
482489	100
555123	100
124226	200
232721	200
834305	200
490514	300
573401	300
275321	300
288475	300

Option 2b: If you are assigning groups, but you do not yet have individual information or you won't need it for your final analysis, your data set should include only group IDs.

OfficeID
100
101
102
103
104
201
202
203
301
302
303
304

Each row represents a single observation (client, counselor, provider, office, and so on). Each observation must have an individual identifier. This is a unique code for each participant who will be assigned to the service recipient group or the comparison group. The identifiers will enable you to determine who will receive the service, allowing you to notify participants. The identifiers will also enable you to combine (merge) data sets.

RECOMMENDED

The Vocational Rehabilitation (VR) Program Evaluation Coach recommends that you include background characteristics and pretest or preservice data, if applicable, in your data set. You will use this information to make sure that the randomly assigned service recipient (intervention) and comparison groups are similar before you introduce the service. If you include these additional variables, your data set will look like Exhibit 2.

Exhibit 2. Summary table

Each row represents a single observation (client, counselor, office, and so on).

Categorical variables must be converted to binary variables with two possible values: 1 instead of yes and 0 instead of no.

CustomerID	OfficeID	Female	Age	Employed_at_ application	Individualized_ education_ program	Race_white	Race_black	Race_other
159508	100	1	16	1	0	1	0	0
694677	100	0	17	1	1	0	1	0
807588	100	1	16	0	1	0	0	1
482489	100	0	18	0	0	0	0	1
555123	100	1	19	1	0	1	0	0
124226	200	0	17	0	1	1	0	0
232721	200	1	17	0	1	0	1	0
834305	200	1	18	1	0	NA	NA	NA
490514	NA	NA	NA	NA	NA	NA	NA	NA
573401	NA	NA	NA	NA	NA	NA	NA	NA

Missing data are replaced with NA so that the VR Program Evaluation Coach doesn't include a value for that observation.

Each variable has its own column.

NOTE: You can work with data in several programs. If you do not have access to statistical software, you can use Microsoft Excel to prepare your data. In the rest of this guide, we have included some tips that will help you manage your data in Excel.

VR Program Evaluation Coach

STEP 1. IDENTIFYING DATA SOURCES

You should use several types of data to create the initial data set. A checklist and description of the data follow. Examples at the end of this guide show what each data set should look like.

- **List of individuals who will be randomly assigned (required).** Compile a list of all potential service recipients. For example, if you plan to randomly select clients to receive the service, you will need an identifier for every client who might be assigned to either the service recipient group or the comparison group. If the clients are grouped under different counselors or offices, you should also include an identifier for the counselor or office. If you plan to randomly select counselors, offices, or providers, you will need an identifier for each. (Remember, larger sample sizes are preferable.)
- **Pretest or preservice data (optional).** If you are testing a change in an outcome, such as an assessment score, over time, you should include a measure of the outcome from before the introduction of the service. The VR Program Evaluation Coach will double-check that the intervention and comparison groups are equivalent on the pre-service measure before giving you the final list of assignments. If the groups are not balanced, the VR Program Evaluation Coach will rerandomize the list of recipients until it achieves balance.
- **Background characteristics (recommended).** Background characteristics provide data on observable traits for each individual. These could include gender, race/ethnicity, age, disability category, socioeconomic status, and more. If you include background traits in your data set, the VR Program Evaluation Coach can make sure your intervention and comparison groups are well balanced before introduction of the service.

CAUTION: Some of these background characteristics can change over time and even change as a result of receiving a service. *Therefore, it is preferable to measure and record all background characteristics before introducing the service.*

STEP 2. PROCESSING YOUR DATA

When you have identified the data elements and data sources, you need to combine all data elements into one **tidy data**[1] set and prepare the variables that will be used for analysis. (We explain below how to do this.) We recommend generating tidy data sets not only because doing so is a requirement for using the VR Program Evaluation Coach and using most statistical software packages, but also because a tidy data set is easy to manipulate, model, and visualize. The data set at the beginning of this guide is an example of a tidy data set. This section will take you through a series of questions that will help you create your own tidy data set.

[1] When you have **tidy data,** you've used a standardized way to structure your data set. In other words, (1) each variable forms a column, (2) each observation forms a row, and (3) each type of observational unit forms a table. If you want to learn more about tidy

A. Is each observation a row and each variable a column (Exhibit 3)?

Exhibit 3. Example: Columns and rows

CustomerID	Treatment	Gender	Employed_at_application
159508	1	F	1
694677	0	M	1
807588	0	F	0
482489	1	M	0

NO: Reorganize your data set so that each row represents one observation. Each variable you are interested in should be its own column.

YES: Continue to B.

B. Do you have one data set that contains all of the variables that you will need?

NO: You will have to merge your existing data sets into one complete data set. The unique identifiers make it easy to merge your data sets. If you are using Excel to manage your data, you can merge data sets by using a VLOOKUP function, (https://support.office.com/en-us/article/VLOOKUP-function-0bbc8083-26fe-4963-8ab8-93a18ad188a1).

CAUTION: Some observations could be present in some data sets but not in others. Therefore, when merging data sets you might introduce some missing data. For example, a client in a data set consisting of applicant information might not be included in another data set and therefore could have missing data for other variables (such as background characteristics) after combining the data sets.

YES: Continue to C.

NOTE: If you are using a statistical software package, you need to make sure that all of the variables are recognized as numeric values and not as string or character values.

C. Are all of the categorical variables that will be used for your analysis numeric (Exhibit 4)?

NO: Convert all of your categorical variables, or variables that include names or labels, into numbers. This might mean that you have to change a variable into a binary, or dummy, variable. A dummy variable uses 1 to indicate yes or that a condition was met and 0 to indicate no or that a condition was not met.

---

data, refer to Wickham, H. "Tidy Data." *Journal of Statistical Software,* vol. 59. no. 10, 2014, pp. 1–23. doi: http://dx.doi.org/10.18637/jss.v059.i10.

For example, if your background characteristics include gender as a variable, you might have "male" or "female," or "M" or "F" in each cell of that column. Instead, you should change the variable name from Gender to Female, and change each cell indicating that the participant is female to 1, and each cell indicating that the participant is male to 0. You can do this for every variable that is non-numeric.

Exhibit 4. Example: Categorical and numeric variables

CustomerID	Gender
159508	F
694677	M
807588	F

→

CustomerID	Female
159508	**1**
694677	**0**
807588	**1**

NOTE: If your categorical variable contains more than two options, such as Race (in which the options are black, white, and other), you will have to create a binary or dummy variable for each option. For example, black would be one column (with 0 representing nonblacks and 1 representing blacks); white would be a second column, and other would be a third column.

YES: Continue to D.

D. Are all missing data coded consistently in your data set?

NO: If you have merged data sets and data are missing, you need to make sure that you are consistently coding the missing data as NA (not available) to be certain that the VR Program Evaluation Coach can analyze your data. You need to be extra careful to ensure that missing data have not been given a numerical designation, such as 0 or 999. These values will get incorporated into the analysis. If you are using your own statistical software, such as SAS or Stata, it will be helpful to code the missing data as a period (.)

YES: Congratulations! You have a tidy data set!

STEP 3. CHECKING THE QUALITY OF YOUR DATA

After constructing your data file and converting your variables, the final step is to check the quality of your data. You can run the following checks to identify potential data issues that warrant additional investigation:

- **Check the minimum and maximum values of variables.** This check can help identify extremely low or high values that are outliers in your distribution or that signal a special missing code that must be converted to a missing value. You might want to check with someone who is familiar with the data to confirm that the value range makes sense.

Note: If you are working in Excel, you can use MIN, (https://support.office.com/en-us/article/MIN-function-61635d12-920f-4ce2-a70f-96f202dcc152) and MAX, (https://support.office.com/en-us/article/MAX-function-e0012414-9ac8-4b34-9a47-73e662c08098) functions to find these values easily.

- **Consider the impact of missing data.** The VR Program Evaluation Coach and some statistical software packages automatically drop observations that contain missing data. You should try to understand why data are missing and how excluding individuals with incomplete data can affect your results.

Note: If you are working in Excel, you can sort and filter your data to view missing values. To determine exactly how many values are missing for a single variable, you can use the COUNTIF function, (https://support.office.com/en-us/article/COUNTIF-function-e0de10c6-f885-4e71-abb4-1f464816df34); to determine how many observations have at least one missing value, you can use a nested COUNTIF with OR function, (https://support.office.com/en-us/article/Use-nested-functions-in-a-formula-9d7c966d-6030-4cd6-a052-478d7d844166).

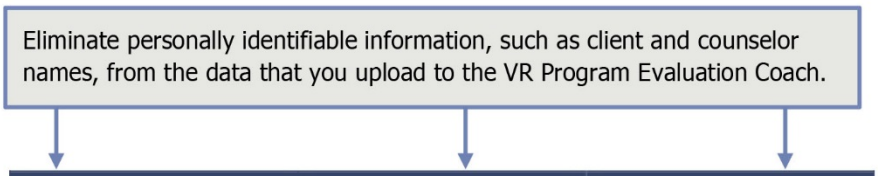
Example data sets

Exhibit 5. Example data set 1a: List of participants

You will need unique identifiers for each participant. If you are randomizing groups of participants (such as under counselors or offices), you must also include a group ID.

Customer_name	CustomerID	Counselor_name	CounselorID	OfficeID
Guillermo Gonzalez	159508	Ken Adams	01	100
Robert Rice	694677	Ken Adams	01	100
Sophia Smith	807588	Josephine Hill	02	100
Patricia Pacheco	482489	Josephine Hill	02	100

Exhibit 6. Example data set 1b: List of participants without personally identifiable information



CustomerID	CounselorID	OfficeID
159508	01	100
694677	01	100
807588	02	100
482489	02	100

Exhibit 7. Example data set 2: Background characteristics (with non-numeric and numeric categorical variables)

CustomerID	OfficeID	Gender	Employed_at_application	Race
159508	100	Female	Yes	White
694677	100	Male	Yes	Black
807588	100	Female	No	Other
482489	100	Male	No	Other
555123	100	Female	Yes	White
124226	200	Male	No	White
232721	200	Female	No	Black
834305	200	Female	Yes	Missing

Background characteristics with non-numeric categorical variables

Background characteristics with numeric categorical variables

CustomerID	OfficeID	Female	Employed_at_application	Race_white	Race_black	Race_other
159508	100	1	1	1	0	0
694677	100	0	1	0	1	0
807588	100	1	0	0	0	1
482489	100	0	0	0	0	1
555123	100	1	1	1	0	0
124226	200	0	0	1	0	0
232721	200	1	0	0	1	0
834305	200	1	1	NA	NA	NA

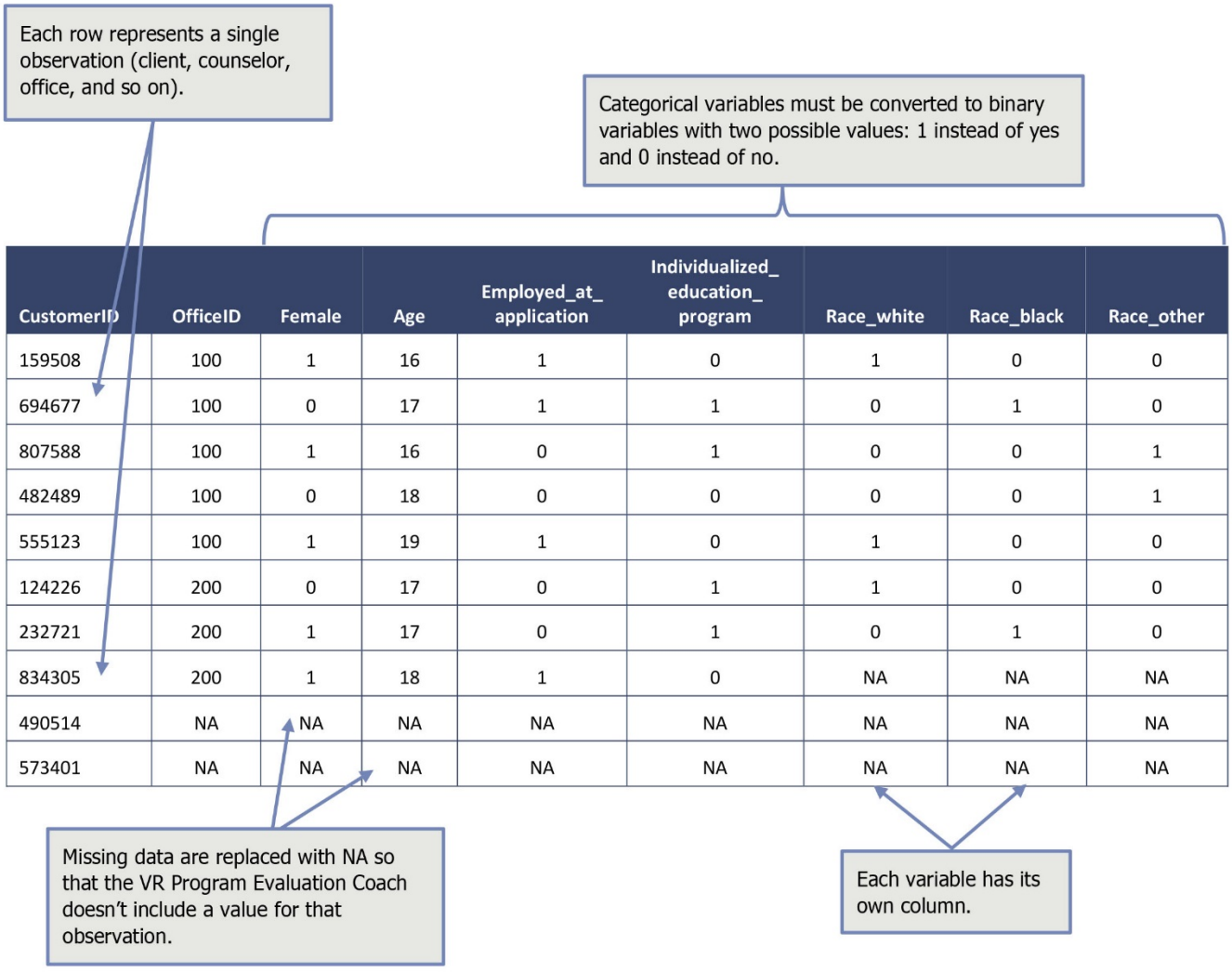
The VR Program Evaluation Coach requires all categorical variables to be converted to binary or dummy variables with values of 0 and 1.

Prepare Your Data for Analysis: Randomized Pilot

INTRODUCTION

To answer your program evaluation question, you will need data from several sources. If you need guidance on processing your data, this document will take you through best practices. At the end, you will have a data file that will look like Exhibit 1 and be ready to upload to the Vocational Rehabilitation (VR) Program Evaluation Coach.

Exhibit 1. Summary table



CustomerID	OfficeID	Female	Age	Employed_at_application	Individualized_education_program	Race_white	Race_black	Race_other
159508	100	1	16	1	0	1	0	0
694677	100	0	17	1	1	0	1	0
807588	100	1	16	0	1	0	0	1
482489	100	0	18	0	0	0	0	1
555123	100	1	19	1	0	1	0	0
124226	200	0	17	0	1	1	0	0
232721	200	1	17	0	1	0	1	0
834305	200	1	18	1	0	NA	NA	NA
490514	NA	NA	NA	NA	NA	NA	NA	NA
573401	NA	NA	NA	NA	NA	NA	NA	NA

NOTE: You can work with data in several programs. If you do not have access to statistical software, you can use Microsoft Excel to prepare your data. We have included some tips throughout this guide that will help you manage your data in Excel.

VR Program Evaluation Coach

STEP 1. IDENTIFYING DATA SOURCES

You will use several types of data in your analysis. A checklist and description of the data you should have on hand follow. At the end of this guide are examples showing what each data set should look like.

CONSIDER: It is important that all of your data are recorded by using **individual identifiers.** These are unique codes for each individual in your data set. The identifiers are what will enable you to combine (merge) data sets. These could be a client or customer ID number, office ID number, counselor ID number, and so on. If you have clients under several counselors or providers under several offices, you should include an ID for both the individual (for example, clients) and group (for example, offices).

- **Treatment status (required).** This variable is in the data set that you downloaded from the VR Program Evaluation Coach after random selection of your service recipients and your comparison group. It indicates which individuals did and did not receive access to the service. Typically, a 1 in this column indicates access to the service (treatment or intervention group) and a 0 indicates no access to the service (comparison group).

CAUTION: You must use the original file with the original treatment status variable, even if someone in your comparison group found a way to receive the service, or someone in your intervention group did not receive the service. The VR Program Evaluation Coach will calculate the effect of random assignment to receive the service on your outcome of interest, regardless of how much exposure participants actually received. If you move individuals from one group to another after randomization, your groups might be unbalanced, and you will not be able to make reliable conclusions about the service's effect.

- **Outcome data (required).** These are data on the outcome you are using to determine the effect of your service, such as rate of exiting with employment.
- **Pretest or preservice data (optional).** If you are testing a change in an outcome, such as an assessment score over time, this is a measure of the outcome before the introduction of the service.
- **Background characteristics (recommended).** Background characteristics provide data on observable traits for each individual. These could include gender, race/ethnicity, age, disability category, socioeconomic status, and more. It is necessary to include background characteristics in your analysis in case a lot of participants drop out of one group. (For instance, if all individuals remain in the service recipient group but 25 percent drop out of the comparison group, you might suspect that the two groups now differ.) Accounting for background characteristics can rebalance the two groups. However, the VR Program Evaluation Coach recommends that you include some background characteristics in your analysis even if your program evaluation went as planned.

CAUTION: Some background characteristics can change over time and even change as a result of receiving a service. *Therefore, it is preferable to measure all background characteristics before introducing the service.*

- **Service delivery data (recommended).** These are data on how the service was implemented, patterns of use over time, and whether the service was delivered as intended. This information will help you interpret your program evaluation findings.

STEP 2. PROCESSING YOUR DATA

After you have identified the data elements and data sources, you need to combine all data elements into one **tidy data**[1] set and prepare the variables that you will use for analysis. (We explain below how to do this.) We recommend that you generate tidy data sets not only because it is a requirement for using the VR Program Evaluation Coach and using most statistical software packages for analyses, but also because a tidy data set is easy to manipulate, model, and visualize. The data set at the beginning of this guide is an example of a tidy data set. This section will take you through a series of questions that will help you create your own.

A. Is each observation a row and each variable a column (Exhibit 2)?

Exhibit 2. Example: Columns and rows

CustomerID	Treatment	Gender	Employed_at_application
159508	1	F	1
694677	0	M	1
807588	0	F	0
482489	1	M	0

NO: Reorganize your data set so that each row represents one observation. Each variable you are interested in should be its own column.

[1] When you have tidy data, you've used a standardized way to structure your data set. In other words, (1) each variable forms a column; (2) each observation forms a row; and (3) each type of observational unit forms a table. If you want to learn more about tidy data, see Wickham, H. "Tidy Data." *Journal of Statistical Software,* vol. 59, no. 10, 2014, pp.1–23. doi: http://dx.doi.org/10.18637/jss.v059.i10.

NOTE: The data set that you downloaded from the VR Program Evaluation Coach after random assignment might look different from the data set you now need for your analysis. The analysis data set should contain one row for each individual or group for which you measured outcomes.

For example, if you randomly assigned high school student clients to receive work-based learning experiences and you want to look at client employment outcomes, you probably have to add only the employment outcomes to the original data set of clients. However, if you randomly assigned counselors to work with high school student clients on work-based learning experiences and you want to look at client employment outcomes, your original data set had one row per counselor, but the data set should now have one row per client. You must include both client and counselor IDs. In addition, client characteristics should be individual to each student, not the average of each counselor.

YES: Continue to B.

B. Do you have one data set that contains all of the variables you will need for analysis?

NO: You will have to merge your existing data sets into one complete data set. This will be easy to do by using the unique identifiers. If you are using Excel to manage your data, you can do this by using a VLOOKUP function, (https://support.office.com/en-us/article/VLOOKUP-function-0bbc8083-26fe-4963-8ab8-93a18ad188a1).

CAUTION: Some observations might be present in some data sets but not in others. Therefore, when merging data sets you could introduce some missing data. For example, a client in a data set consisting of applicant information might not be included in another data set and therefore could have missing data for other variables (such as background characteristics) after combining the data sets.

YES: Continue to C.

NOTE: If you are using a statistical software package, you need to make sure that all of the variables are recognized as numeric values and not as string or character values.

C. Are all of the categorical variables that your analysis will use numeric (Exhibit 3)?

NO: Convert all of your categorical variables, or variables that include names or labels, into numbers. This might mean that you have to change a variable into a binary or dummy variable. A dummy variable uses 1 to indicate yes or that a condition was met and 0 to indicate no or that a condition was not met.

For example, if your background characteristics include gender as a variable, you might have “male” or “female,” or “M” or “F,” in each cell of that column. Instead, you should change the variable name from Gender to Female, and change each cell indicating that the

participant is female to 1, and each cell indicating that the participant is male to 0. You can do this for every variable that is non-numeric.

Exhibit 3. Example: Categorical and numeric variables

CustomerID	Gender
159508	F
694677	M
807588	F

CustomerID	Female
159508	**1**
694677	**0**
807588	**1**

NOTE: If your categorical variable contains more than two options, such as Race (in which the options are black, white, and other), you will have to create a binary or dummy variable for each option. For example, black would be one column (with 0 representing nonblacks and 1 representing blacks); white would be a second column; and other would be a third column.

YES: Continue to D.

D. Are all missing data coded as NA in your data set?

NO: If you have merged data sets and there are missing data, you need to make sure that you are consistently coding the missing data as NA to be certain that the VR Program Evaluation Coach can analyze your data. You need to be extra careful to ensure that missing data have not been given a number designation, such as 0 or 999. These values will get incorporated into the analysis. If you are using your own statistical software, such as SAS or Stata, it will be helpful to code the missing data as a period (.)

YES: Congratulations! You have a tidy data set!

STEP 3. CHECKING THE QUALITY OF YOUR DATA

After constructing your data file and converting your variables, the final step is to check the quality of your data. You can run the following checks to identify potential data issues that warrant additional investigation:

- **Check the minimum and maximum values of variables.** This check can help identify extremely low or high values that are outliers in your distribution or that might signal a special missing code that has to be converted to a missing value. You might want to check with someone who is familiar with the data to confirm that the value range makes sense.

Note: If you are working in Excel, you can use MIN, (https://support.office.com/en-us/article/MIN-function-61635d12-920f-4ce2-a70f-96f202dcc152) and MAX, (https://support.office.com/en-us/article/MAX-function-e0012414-9ac8-4b34-9a47-73e662c08098) functions to find these values easily.

- **Consider the impact of missing data.** The VR Program Evaluation Coach and some statistical software packages automatically drop observations that contain missing data. You should try to understand why data are missing and how excluding individuals with incomplete data can affect your results.

Note: If you are working in Excel, you can sort and filter your data to view missing values. To determine exactly how many values are missing for a single variable, you can use the COUNTIF function, (https://support.office.com/en-us/article/COUNTIF-function-e0de10c6-f885-4e71-abb4-1f464816df34); to determine how many observations have at least one missing value, you can use a nested COUNTIF with OR function, (https://support.office.com/en-us/article/Use-nested-functions-in-a-formula-9d7c966d-6030-4cd6-a052-478d7d844166).

Example data sets

Exhibit 4. Example data set 1: Treatment status and outcome data

CustomerID	Treatment	Employed_at_exit
159508	1	0
694677	0	0
807588	0	1
482489	1	0

Exhibit 5. Example data set 2: Background characteristics (with non-numeric and numeric categorical variables)

CustomerID	OfficeID	Gender	Employed_at_application	Race
159508	100	Female	Yes	White
694677	100	Male	Yes	Black
807588	100	Female	No	Other
482489	100	Male	No	Other
555123	100	Female	Yes	White
124226	200	Male	No	White
232721	200	Female	No	Black
834305	200	Female	Yes	Missing

Background characteristics with non-numeric categorical variables

Background characteristics with numeric categorical variables

CustomerID	OfficeID	Female	Employed_at_application	Race_white	Race_black	Race_other
159508	100	1	1	1	0	0
694677	100	0	1	0	1	0
807588	100	1	0	0	0	1
482489	100	0	0	0	0	1
555123	100	1	1	1	0	0
124226	200	0	0	1	0	0
232721	200	1	0	0	1	0
834305	200	1	1	NA	NA	NA

The VR Program Evaluation Coach requires all categorical variables to be converted to binary or dummy variables with values of 0 and 1.

Guide to Using Matched Comparison Approach

Matching Overview

When random assignment is not possible, you may use matching to create a good comparison group to help you determine what service works.

THE EVALUATION CHALLENGE

You want to test whether a service change is effective, but it is impossible to observe simultaneously what happens when an individual receives or doesn't receive that service. If you introduce a service and watch what happens, you could notice improvements in, for example, rates of exiting with employment. However, you *cannot* assume that the service *caused* improved employment rates. Many other factors (including client work experience, other client characteristics, other VR programs, external factors such as an improving economy, and so on) could have contributed to the increases.

To overcome these challenges, it is important to compare a group of service recipients to a group of nonrecipients, on the assumption that the only real difference between the groups is whether they are receiving the service. However, comparing service recipients and nonrecipients brings an additional set of challenges. When we make comparisons without trying to ensure similarities between groups, it is possible that those who receive the service differ in any number of ways from those who do not receive the service. For example, clients with work experience might be more motivated to receive a new job placement service, but they may also be more likely to exit with employment by virtue of that experience. This might cause you to confuse the effect of the service with the effect of employment experience (either could cause service recipients to outperform nonrecipients). Such differences can make an ineffective service look effective, or vice versa.

MATCHED COMPARISON DESIGN

Solution. You can match service recipients to similar nonrecipients by using pretest measures (if applicable) and background characteristics. Once you have created two similar groups, you are comparing apples to apples—the only observed difference between service recipients and nonrecipients is their exposure to the service (though there might be unobserved differences). Then, if you see differences in outcomes (such as rates of exiting with employment), you may be confident that the new service is producing the desired outcome.

How it works. Suppose that you have a student, Jane (shown in Exhibit 1), who enrolled in a new work-based learning experience program. After she completes the program, Jane eventually graduates from high school and exits successfully from VR (that is, she finds a job). To know if the program is having a positive effect on exiting successfully from the program, you would like to be able to observe a Jane in a parallel universe. "Parallel Jane" is exactly the same as Jane but has no access to the new work-based learning experiences program. If Parallel Jane did not exit successfully, you could conclude that the new program had the desired outcome for Jane.

Exhibit 1. Jane's characteristics and outcomes

TREATMENT	JANE
	Age: 17
	Grade: Junior
	Gender: Female
	Graduated high school: Yes
	Exited successfully: Yes

With matching. Your initial comparison, without matching, compares Jane with the average youth clients across the state. Matching attempts to find someone as close as possible to Parallel Jane. When the program is successful, you may conclude that any differences in outcomes are attributable to the new program, not to other factors.

You may compare Jane with one of the four potential comparison students shown in Exhibit 2. The better the match, the more confident you may be in your conclusion that the new work-based learning experiences program is producing the desired impact. *With whom should you compare Jane?*

Exhibit 2. Potential comparisons

POTENTIAL COMPARISON	JOHN	JILL	JENNY	JODY
	Age: 18	Age: 18	Age: 17	Age: 16
	Grade: Senior	Grade: Senior	Grade: Junior	Grade: Sophomore
	Gender: Male	Gender: Female	Gender: Female	Gender: Female
	Graduated high school: Yes	Graduated high school: No	Graduated high school: Yes	Graduated high school: Yes
	Exited successfully: No	Exited successfully: No	Exited successfully: Yes	Exited successfully: Yes

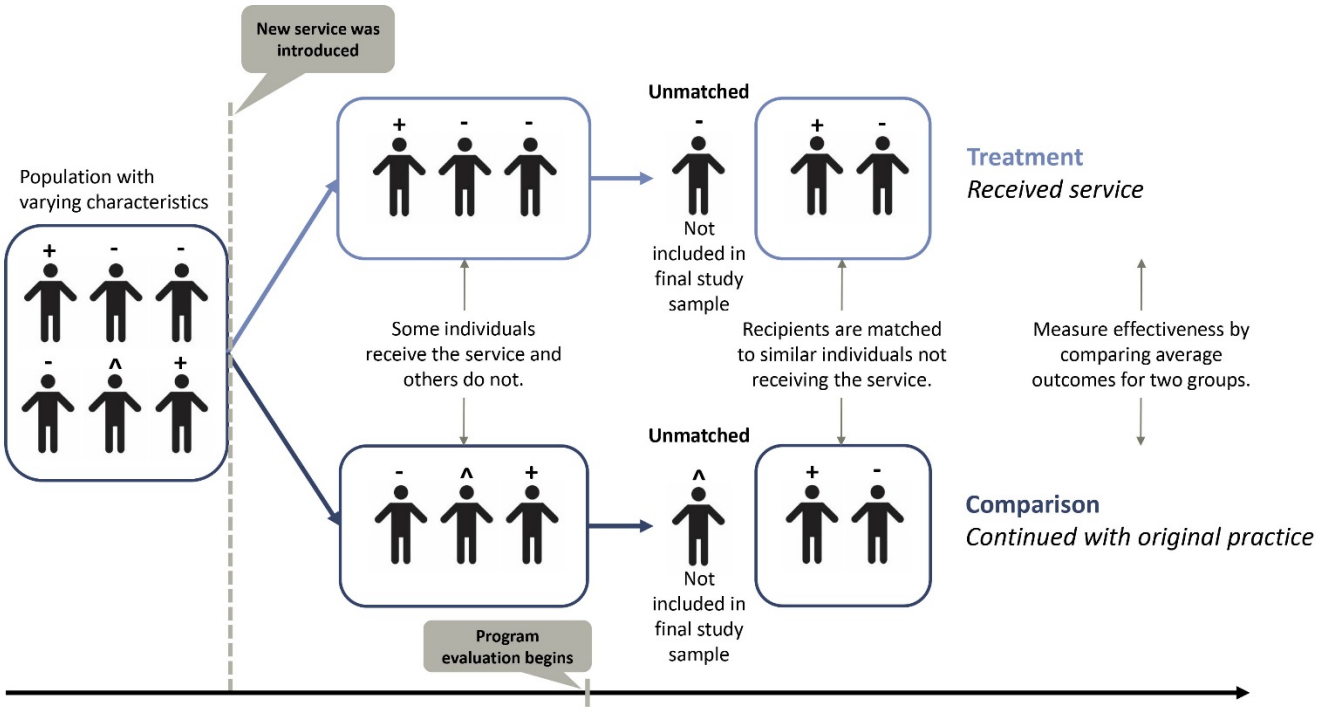
We want to match Jane with someone with similar characteristics during the same observation period. Matching on a preservice outcome (such as a test score) is the best approach for the matching technique to work, but VR agencies often do not have a good preservice outcome measure related to VR service provision. In such situations, you should select one or more other relevant baseline measures or characteristics to use for matching.

In practice, even if you do have a good preservice outcome measure, you will want to use more than one observed characteristic to find matches for individuals receiving a service. To do so, the VR Program Evaluation Coach matching tool uses a statistical technique called nearest neighbor matching. In the example above, we would match Jane to Jenny, who has similar background characteristics as Jane. Then, we use other statistical techniques to compare their rates of exiting successfully and determine if the new work-based learning experiences program is or is not having the desired outcome for Jane and others in the program.

Note: If the service or program is targeted to a highly specific group (or a specific type of individual is likely to receive the service), the VR Program Evaluation Coach is less able to identify a good matched comparison.

For your program evaluation, the matching process will resemble the outcome in Exhibit 3.

Exhibit 3. The matching process



BASELINE EQUIVALENCE

An important question to ask is, Was the matching successful? A common way to answer the question is to compare the average values of the groups' background characteristics after matching. For characteristics such as assessment results, earlier years' employment rates, or demographic

characteristics, we can quantify the difference between the two groups by using a measure called an **effect size**. Researchers use an effect size to measure different characteristics by using the same yardstick. Effect size is calculated by dividing the difference in means between the two groups by the standard deviation of the entire sample. The VR Program Evaluation Coach's matching dashboard automatically calculates the difference (measured as an effect size) between the recipient group average and the matched nonrecipient group average for any variables you specify. The levels of balance follow the U.S. Department of Education's What Works Clearinghouse (WWC) standards defined in Exhibit 4. Meeting the WWC standard for baseline equivalence helps bolster confidence that any effects are the result of the new service or program.

Exhibit 4. WWC standards for baseline equivalence

Absolute value of difference between groups	WWC conclusion on baseline equivalence
Effect size ≤ 0.05	Satisfies baseline equivalence requirement
0.05 < effect size ≤ 0.25	Requires statistical adjustment
0.25 < effect size	Does not satisfy baseline equivalence requirement

If the VR Program Evaluation Coach's matching dashboard finds that your groups differ in an important baseline characteristic greater than 0.25 effect size units, it will prevent you from advancing to the next step until you develop a better a match.

Caution: The amount of confidence that you have in the results of a matched comparison analysis is based on the similarity between the groups of recipients and matched nonrecipients. Two important notes follow from this:

1. The use of observed characteristics to create a matched comparison does not necessarily yield two similar groups. It's important to think about the variables you include. Too few variables can lead to groups that aren't similar, and too many variables, particularly variables that aren't important for the outcome, make it too difficult to match similar individuals. Focus on the variables that you think are important to the outcome or the likelihood of receiving the service.
2. Even if you have a large set of observed characteristics, a matched comparison analysis cannot eliminate the possibility that individuals receiving the service differ in some unobserved way from those not receiving it. For example, youth who receive the work-based learning experiences program might have parents who are more involved and advocate for additional attention or VR resources as compared to the parents of youth who are not receiving the program. Or counselors who devote extra effort to implementing a new service might work especially hard to enhance other aspects of their work with clients. If so, the measured effect of the intervention could be inaccurate, even if the groups appear to be well matched on observed characteristics.

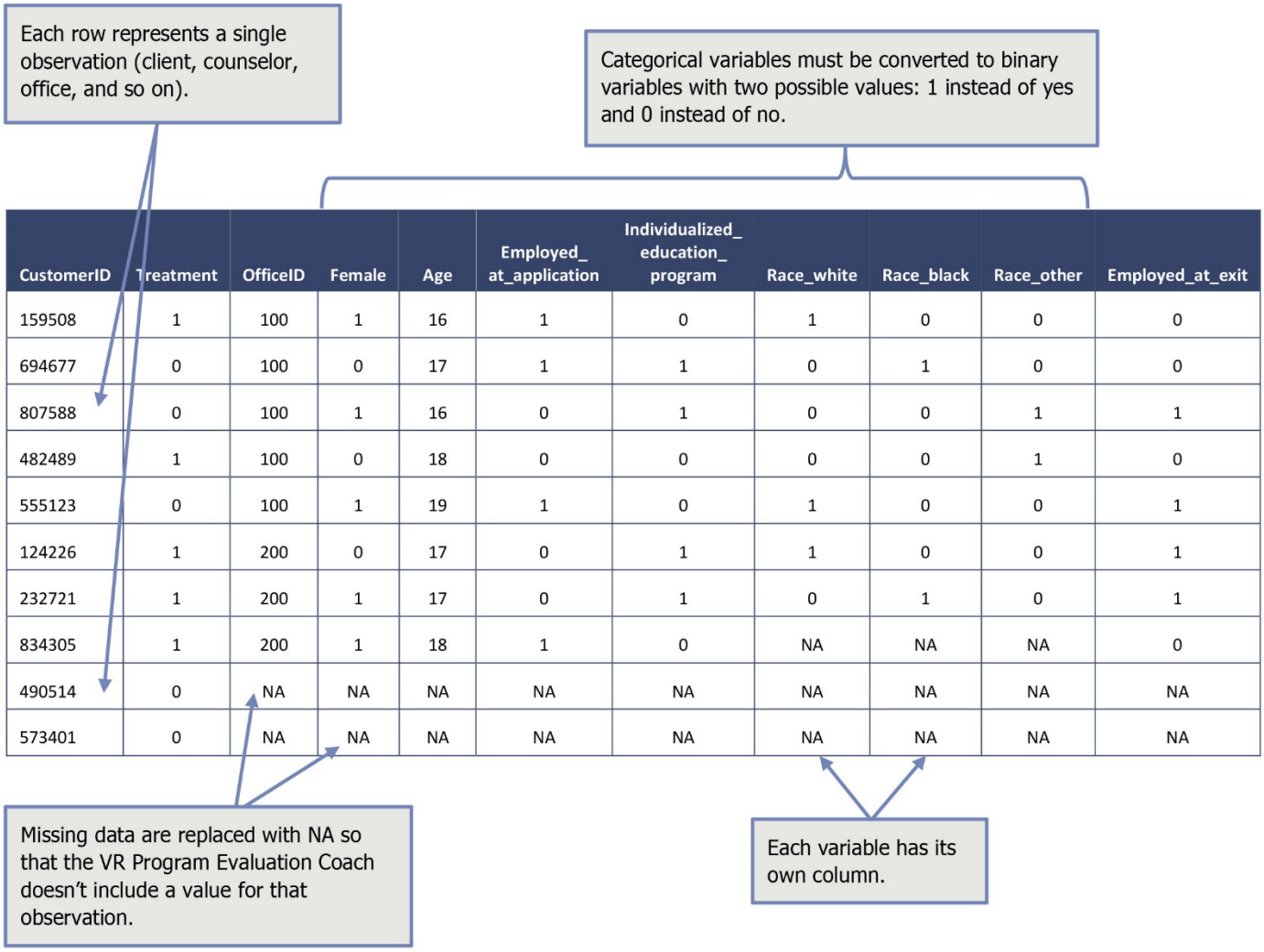
The only way to remove those potential differences between the two groups is to use random assignment to select the treatment and comparison groups. When properly conducted, random assignment ensures that the two groups are similar in both observed and unobserved characteristics.

Prepare Your Data for Analysis: Matched Comparison Design

INTRODUCTION

To answer your program evaluation question, you will need data from several sources. If you need guidance on processing your data, this document will take you through best practices. At the end, you will have a data file that will look like Exhibit 1 and be ready to upload to the VR Program Evaluation Coach.

Exhibit 1. Summary table



CustomerID	Treatment	OfficeID	Female	Age	Employed_ at_application	Individualized_ education_ program	Race_white	Race_black	Race_other	Employed_at_exit
159508	1	100	1	16	1	0	1	0	0	0
694677	0	100	0	17	1	1	0	1	0	0
807588	0	100	1	16	0	1	0	0	1	1
482489	1	100	0	18	0	0	0	0	1	0
555123	0	100	1	19	1	0	1	0	0	1
124226	1	200	0	17	0	1	1	0	0	1
232721	1	200	1	17	0	1	0	1	0	1
834305	1	200	1	18	1	0	NA	NA	NA	0
490514	0	NA	NA	NA	NA	NA	NA	NA	NA	NA
573401	0	NA	NA	NA	NA	NA	NA	NA	NA	NA

NOTE: You can work with data in several programs. If you do not have access to statistical software, you can use Microsoft Excel to prepare your data. We have included some tips throughout this guide that will help you manage your data in Excel.

VR Program Evaluation Coach

STEP 1. IDENTIFYING DATA SOURCES

You will use several types of data in your analysis. Below is a checklist and description of the data you should have on hand. At the end of this guide are examples showing what each data set should look like.

CONSIDER: It is important that all of your data are recorded by using **individual identifiers**. These are unique codes for each individual in your data set. The identifiers are what will enable you to combine (merge) data sets. These could be a client or customer ID number, office ID number, counselor ID number, and so on. If you have clients under several counselors or providers under several offices, you should include an ID for both the individual (for example, clients) and the group (for example, offices).

- **Treatment status (required).** This is the data element that indicates which individuals did and did not receive access to the service. Typically, a 1 in this column indicates access to the service (treatment or intervention group) and a 0 indicates no access to the service (comparison group).
- **Outcome data (required).** These are data on the outcome you are using to determine the effect of your service, such as rate of exiting with employment.
- **Pretest or preservice data (optional).** If you are testing a change in an outcome, such as an assessment score over time, this is a measure of the outcome before the introduction of the service.
- **Background characteristics (recommended).** Background characteristics provide data on observable traits for each individual. These could include gender, race/ethnicity, age, disability category, socioeconomic status, and more. Background traits are particularly important for matched comparison program evaluations because you can use them to make sure your intervention and comparison groups are well matched (balanced).

CAUTION: Some background characteristics can change over time and even change as a result of receiving a service. *Therefore, it is preferable to measure all background characteristics before introducing the service.*

- **Service delivery data (recommended).** These are data on how the service was implemented, patterns of use over time, and whether the service was delivered as intended. This information will help you interpret your program evaluation findings. Service delivery information can be used to assign individuals to intervention and comparison groups.

STEP 2. PROCESSING YOUR DATA

After you have identified the data elements and data sources, you need to combine all data elements into one **tidy data**[1] set and prepare the variables that will be used for analysis. (We explain below how to do this.) We recommend that you generate tidy data sets not only because it is a requirement for using the VR Program Evaluation Coach and using most statistical software packages for analyses, but also because a tidy data set is easy to manipulate, model, and visualize. The data set at the beginning of this guide is an example of a tidy data set. This section will take you through a series of questions that will help you create your own.

A. Is each observation a row and each variable a column (Exhibit 2)?

Exhibit 2. Example: Columns and rows

CustomerID	Treatment	Gender	Employed_at_application
159508	1	F	1
694677	0	M	1
807588	0	F	0
482489	1	M	0

NO: Reorganize your data set so that each row represents one observation. Each variable you are interested in should be its own column.

YES: Continue to B.

B. Do you have one data set that contains all of the variables you will need for analysis?

NO: You will have to merge your existing data sets into one complete data set. This will be easy to do by using the unique identifiers. If you are using Excel to manage your data, you can do this by using a VLOOKUP function, (https://support.office.com/en-us/article/VLOOKUP-function-0bbc8083-26fe-4963-8ab8-93a18ad188a1).

[1] **When you have tidy data,** you've used a standardized way to structure your data set. In other words, (1) each variable forms a column, (2) each observation forms a row, and (3) each type of observational unit forms a table. To learn more about tidy data, see Wickham, H. "Tidy Data." *Journal of Statistical Software*, vol. 59, no. 10, 2014, pp.1–23. doi: http://dx.doi.org/10.18637/jss.v059.i10.

CAUTION: Some observations might be present in some data sets but not in others. Therefore, when merging data sets, you might introduce some missing data. For example, a client in a data set consisting of applicant information might not be included in another data set and therefore could have missing data for other variables (such as treatment status and background characteristics) after combining the data sets.

YES: Continue to C.

NOTE: If you are using a statistical software package, you will need to make sure that all of the variables are recognized as numeric values and not as string or character values.

C. Are all of the categorical variables that will be used for your analysis numeric (Exhibit 3)?

NO: Convert all of your categorical variables, or variables that include names or labels, into numbers. This might mean you have to change a variable into a binary or dummy variable. A dummy variable uses 1 to indicate yes or that a condition was met and 0 to indicate no or that a condition was not met.

For example, if your background characteristics include gender as a variable, you might have "male" or "female," or "M" or "F," in each cell of that column. Instead, you should change the variable name from Gender to Female, and change each cell indicating that the participant is female to 1, and each cell indicating that the participant is male to 0. You can do this for every variable that is non-numeric.

Exhibit 3. Example: Categorical and numeric variables

CustomerID	Gender
159508	F
694677	M
807588	F

CustomerID	Female
159508	**1**
694677	**0**
807588	**1**

NOTE: If your categorical variable contains more than two options, such as Race (in which the options are black, white, and other), you will have to create a binary or dummy variable for each option. For example, black would be one column (with 0 representing nonblacks and 1 representing blacks), white would be a second column, and other would be a third column.

YES: Continue to D.

D. Are all missing data coded as NA in your data set?

NO: If you have merged data sets and data are missing, you need to make sure that you are consistently coding the missing data as NA to be certain that the VR Program Evaluation Coach can analyze your data. You need to be extra careful to ensure that missing data have not been given a number designation, such as 0 or 999. These values will get incorporated

into the analysis. If you are using your own statistical software, such as SAS or Stata, it will be helpful to code the missing data as a period (.)

YES: Congratulations! You have a tidy data set!

STEP 3. CHECKING THE QUALITY OF YOUR DATA

After constructing your data file and converting your variables, the final step is to check the quality of your data. You can run the following checks to identify potential data issues that warrant additional investigation:

- **Check the minimum and maximum values of variables.** This check helps identify extremely low or high values that are outliers in your distribution or that can signal a special missing code that must be converted to a missing value. You might want to check with someone who is familiar with the data to confirm that the value range makes sense.

Note: If you are working in Excel, you can use MIN, (https://support.office.com/en-us/article/MIN-function-61635d12-920f-4ce2-a70f-96f202dcc152) and MAX, (https://support.office.com/en-us/article/MAX-function-e0012414-9ac8-4b34-9a47-73e662c08098) functions to find these values easily.

- **Consider the impact of missing data.** The VR Program Evaluation Coach and some statistical software packages automatically drop observations that contain missing data. You should try to understand why data are missing and how excluding individuals with incomplete data can affect your results.

Note: If you are working in Excel, you can sort and filter your data to view missing values. To determine exactly how many values are missing for a single variable, you can use the COUNTIF function, (https://support.office.com/en-us/article/COUNTIF-function-e0de10c6-f885-4e71-abb4-1f464816df34); to determine how many observations have at least one missing value, you can use a nested COUNTIF with OR function, (https://support.office.com/en-us/article/Use-nested-functions-in-a-formula-9d7c966d-6030-4cd6-a052-478d7d844166).

Example data sets

Exhibit 4. Example data set 1: Treatment status and outcome data

CustomerID	Treatment	Employed_at_exit
159508	1	0
694677	0	0
807588	0	1
482489	1	0

Exhibit 5. Example data set 2: Background characteristics (with non-numeric categorical variables)

CustomerID	OfficeID	Gender	Employed_at_application	Race
159508	100	Female	Yes	White
694677	100	Male	Yes	Black
807588	100	Female	No	Other
482489	100	Male	No	Other
555123	100	Female	Yes	White
124226	200	Male	No	White
232721	200	Female	No	Black
834305	200	Female	Yes	Missing

Background characteristics with non-numeric categorical variables

Background characteristics with numeric categorical variables

CustomerID	OfficeID	Female	Employed_at_application	Race_white	Race_black	Race_other
159508	100	1	1	1	0	0
694677	100	0	1	0	1	0
807588	100	1	0	0	0	1
482489	100	0	0	0	0	1
555123	100	1	1	1	0	0
124226	200	0	0	1	0	0
232721	200	1	0	0	1	0
834305	200	1	1	NA	NA	NA

The VR Program Evaluation Coach requires all categorical variables to be converted to binary or dummy variables with values of 0 and 1.

Appendix A:
Technical Appendices

VR Program Evaluation Coach

Random Assignment Technical Appendix

This document provides a technical overview of how the Vocational Rehabilitation (VR) Program Evaluation Coach random assignment tool functions. For a brief introduction to randomized pilots as a research design, refer to our "Random Assignment Overview" guide available at https://vrevalcoach.com/static/pdf/Random_Assignment_Overview.pdf.

Random assignment attempts to create two groups that are as similar as possible in both observed and unobserved characteristics. You may think of the random assignment process as the process of flipping a coin to assign individuals to either pilot the service or not to the pilot service. Using a fair coin, each individual would have a 50 percent chance of selection into the group piloting the service. Therefore, if you have a sufficiently large pool of potential recipients, the two groups should have the same observed and unobserved characteristics. You may think of the VR Program Evaluation Coach random assignment tool as a way to flip a virtual coin many times, very quickly. In addition, anyone may access the code to see that the coin is truly random. The chance of being assigned to the treatment group may differ from 50 percent to provide a smaller, or larger, fraction of the population for initially testing the service.

WHAT DOES IT MEAN TO HAVE A SUCCESSFUL RANDOM ASSIGNMENT?

A successful random assignment process produces treatment and comparison groups that are very similar. When the groups are large, random assignment is likely to succeed in creating two similar groups. For smaller samples, it is possible that the two groups differ in important ways even though a random process was used to create them. The VR Program Evaluation Coach uses the baseline equivalence standard, (https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_brief_baseline_080715.pdf) from the What Works Clearinghouse to assess whether the two groups are sufficiently similar. We encourage you to use some pre-intervention measure that is related to the outcome to check for baseline balance, but use of such a measure is not required. The VR Program Evaluation Coach considers that random assignment worked as expected if the differences on all the variables selected to control for balance are less than 0.25 standard deviations; otherwise, the VR Program Evaluation Coach attempts the random assignment process again. The VR Program Evaluation Coach assumes that random assignment worked as expected if no variables are specified to check for baseline equivalence.

HOW DOES THE TOOL WORK?

The VR Program Evaluation Coach uses R, a free software package for statistical computing. In particular, the VR Program Evaluation Coach uses the base R function sample.int, (https://stat.ethz.ch/R-manual/R-devel/library/base/html/sample.html) to perform the random assignment. The VR Program Evaluation Coach first performs several checks on the data to confirm

that there are no data problems that would cause an error with the random assignment code. Then, the VR Program Evaluation Coach randomly assigns observations to the intervention or comparison group and, if baseline variables were specified, checks that random assignment worked as expected. The rest of this document describes the specific data checks and the random assignment algorithm.

Step 1: Check for data problems

The VR Program Evaluation Coach performs the following checks to verify that the VR Program Evaluation Coach has the inputs required to perform random assignment:

1. Data are NULL, not readable, or have 0 observations.
2. Variable specifying the randomization unit ID is not specified or does not exist.
3. There is no indication as to whether groups should be determined by either a set number or percentage.
4. Percentage was selected for assignment, and the number of desired recipients is less than or equal to 0 or greater than or equal to 100.
5. Number was selected for assignment, and the number of desired recipients is less than or equal to 0 or greater than or equal to the number of records in the data.
6. Baseline variables are specified, and not all exist in the data.
7. Baseline variables are specified, and one or more are not numeric.
8. A variable named Treatment already exists in the data.

If any test fails, the VR Program Evaluation Coach prints a message to help the user identify the problem. If the VR Program Evaluation Coach detects no problems, it proceeds to Step 2.

Step 2: Randomize and check for balance

The VR Program Evaluation Coach performs the random assignment by using a loop that stops either when random assignment is successful or after 10 failed attempts. Random assignment is deemed successful when all variables selected for the baseline equivalence checks are balanced across the two groups. If the process was successful, the user will be able to download the data and move to the next step in the VR Program Evaluation Coach. Otherwise, the VR Program Evaluation Coach randomizes again with a new random assignment seed. If the VR Program Evaluation Coach fails to produce a successful random assignment 10 times, it informs the user that the process did not work as expected. The following steps describe the loop:

1. The VR Program Evaluation Coach randomly selects a random assignment seed.
2. Using sample.int, several observations are chosen to be in the treatment group. The observations represent either a fixed proportion of the sample (rounded up to the nearest whole observation) or a fixed number of observations.

3. The VR Program Evaluation Coach checks if random assignment was successful.
4. If the process is successful, a file will be available to download with the original data plus one variable indicating whether the observation belongs to the treatment or comparison group.
5. If the process is not successful:
    a. If the loop has run fewer than 10 times, a new seed will be chosen and the loop will run again.
    b. If the loop has run 10 times, a message will explain to the user that the random assignment was not successful.

The code for the VR Program Evaluation Coach is open source under the General Public License Version 3 and will be available soon on our github repository.

VR Evaluation Coach

Matching Technical Appendix

This document provides a technical overview of how the Vocational Rehabilitation (VR) Program Evaluation Coach matching tool functions. For a brief introduction to matching as a research design, you may refer to our "Matching Overview" guide available at https://vrevalcoach.com/static/pdf/Matching_Overview.pdf.

Matching attempts to group individuals (or groups of individuals) who are receiving a new service or program with other individuals who are not receiving the service but who have similar characteristics. The ideal match would be two individuals identical on every observed and unobserved characteristic except for receipt of the service. In practice, it is very unlikely that two individuals will be identical on all observed characteristics, so you must determine which observable characteristics you will match on according to how likely those characteristics are to affect the outcome of interest.

Matching on only one variable is conceptually simple. Suppose that you have a set of high school students receiving a new job readiness program and another set of students without access to the program. You are trying to create a matched data set by using only baseline characteristics. A simple algorithm would look like the following:

1. Define a distance measure. For example, the absolute distance between pretest job readiness measures of 240 and 265 is: |240 - 265|= 25
2. For a student in the new program, calculate the distance between her pretest measure of job readiness and the pretest measure of each of the students not receiving the new program.
3. Match the student in the new program to the student not receiving the new program with the minimum distance.
4. Repeat Steps 2 through 4 for all students receiving the new program.

It is more complicated to define distance when you are trying to match individuals (or groups of individuals) by using more than one characteristic. When you use more than one matching variable, you need to summarize the information into a single distance number. For example, suppose that you want to include an indicator for age and another for economic disadvantage. In this case, we can create a measure that summarizes all the information and match based on the summary—even though the scales differ—by calculating the probability that an individual with a given set of observable characteristics receives the new program. The goal is that, after matching, individuals in the treatment group will be matched to individuals in the comparison group who appear the same on observable key characteristics.

HOW DOES THE TOOL WORK?

The VR Program Evaluation Coach uses R, a free software package for statistical computing. The VR Program Evaluation Coach first performs several checks on the data to confirm that there are no data problems that would cause an error with the matching code. Then, the VR Program Evaluation Coach creates a matched data set by using the R package MatchIt, (https://cran.r-project.org/web/packages/MatchIt/index.html). The rest of this document describes the specific data checks and randomization algorithm.

Step 1: Check for data issues

The VR Program Evaluation Coach performs the following checks to verify that the VR Program Evaluation Coach has the inputs needed to assign the individuals or groups of individuals randomly:

1. Data are NULL, not readable, or have 0 observations.
2. No treatment variable specified.
3. No matching variables specified.
4. One or more matching variables are not found in the data.
5. Treatment variable has values other than 0, 1, and NA.
6. One or more matching variables are not numeric.

If any test fails, the VR Program Evaluation Coach prints a message to help the user identify the problem. If the VR Program Evaluation Coach detects no problems, it proceeds to Step 2.

Step 2: Create a matched data set

To create a matched data set, the VR Program Evaluation Coach uses one-to-one nearest neighbor matching (Rubin and Thomas 1996). The VR Program Evaluation Coach selects a comparison individual with the smallest distance for each individual receiving the service.

After the VR Program Evaluation Coach has selected matches for each individual receiving the service, it assesses whether the two groups—the treatment group and the matched comparison group—are similar on the key characteristics. The VR Program Evaluation Coach uses the baseline equivalence standard from the What Works Clearinghouse, (https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_brief_baseline_080715.pdf) to conduct the assessment. We consider the matched group to be a valid comparison group if the baseline differences on all matching variables are less than 0.25 standard deviations.

If the VR Program Evaluation Coach fails to generate a valid comparison group, it defines a maximum distance, known as a caliper, and repeats the matching algorithm. If the distance between the individual receiving the service and the nearest individual not receiving it is larger than the caliper, the VR Program Evaluation Coach drops the individual receiving the service. If the VR Program Evaluation

Coach cannot generate a valid comparison group with the caliper, it uses a smaller caliper. The VR Program Evaluation Coach starts by using no caliper and then moves from 1.00 standard deviation of the distance measure to 0.25 standard deviations in 0.25 intervals. Using a caliper can mean that some recipients of the service do not receive a match. If this is the case, you should be cautious when interpreting the results because you are measuring the effect only for the individuals with a good match. The algorithm may be summarized as follows:

1. The VR Program Evaluation Coach attempts to match treatment and comparison observations by using nearest neighbor matching.
2. If matching in Step 1 successfully produces balanced samples, the matched groups are returned as a downloadable file, and the screen shows relevant sample sizes and graphic displays of balance.
3. If matching in Step 1 was not successful, the VR Program Evaluation Coach attempts to match by using caliper matching, with an initial caliper value of 1.00 and decreasing by 0.25 until matching produces balanced samples or the caliper reaches 0.
    a. If matching was successful, a downloadable file is available as in Step 2, with the same sample sizes and graphics shown on screen.
    b. If matching was not successful for any caliper value, balance graphics are shown, but there is no file to download.

The code for the VR Program Evaluation Coach is open source under the General Public License Version 3 and will be available soon on our github repository.

Which variables should I include for matching?

You should include some baseline, or pre-intervention, measurement that is related to your outcome of interest if it is available. In the case of the new job readiness program, you could include a pretest measure of job readiness as one of your matching variables. However, not all VR services will have an appropriate pre-service outcome measure, which is acceptable, as long as you have at least one relevant baseline characteristic to use for matching. There are also cases in which you might want to use additional characteristics for matching in order to improve precision.

In our example, if the number of students receiving and not receiving the new program is markedly different (one group is substantially larger than the other), the literature recommends including:

- Variables that affect the likelihood of receiving the new program
- Variables that affect the outcome you want to study (for example, job readiness), controlling for receipt of the new program (Rubin and Thomas 1996; Heckman et al. 1998; Ho et al. 2011)

If you do not have substantially more individuals not receiving the new service than the number of individuals receiving the service, you **should not** include all available characteristics (Ho et. al. 2011). The intuition is that, as you increase the number of characteristics used for matching, it might become

too difficult to find a good match. For example, think about searching for a new apartment: if you include too many characteristics (location, bedrooms, laundry, size, and so on), your search could end up too narrow and you won't find any suitable apartments.

Should matching variables also be included as covariates in the analysis?

YES! Matching is seldom perfect—some imbalance often remains on at least some variables. Regression adjustment for the matching variables can mitigate any remaining imbalance.

REFERENCES

Heckman, J., H. Ichimura, J. Smith, and P. Todd. "Characterizing Selection Bias Using Experimental Data." No. w6699. Cambridge, MA: National Bureau of Economic Research, 1998.

Ho, Daniel, Kosuke Imai, Gary King, and Elizabeth Stuart. "Matchit: Nonparametric Preprocessing for Parametric Causal Inference." *Journal of Statistical Software*, vol. 42, no. 8, 2011, pp. 1-28.

Rubin, Donald B., and Neal Thomas. "Matching Using Estimated Propensity Scores: Relating Theory to Practice." *Biometrics,* vol. 51, no. 2, 1996, pp. 249–264.

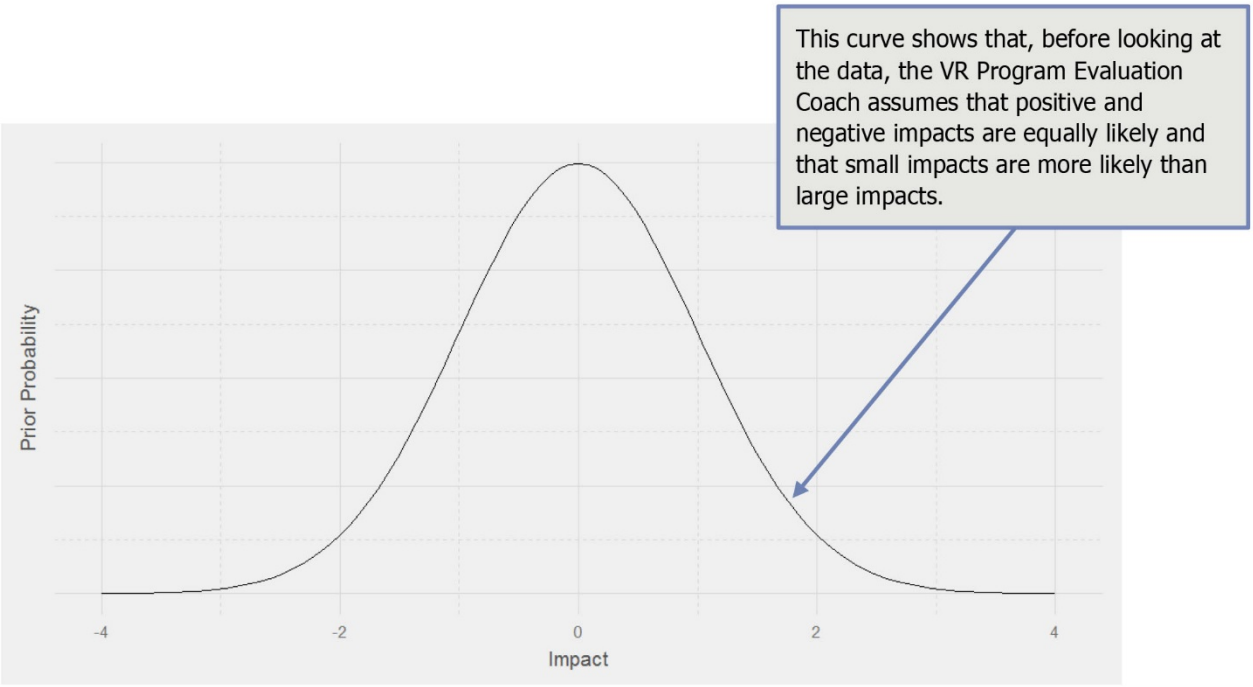
VR Program Evaluation Coach

Impact Estimation Technical Appendix

OVERVIEW

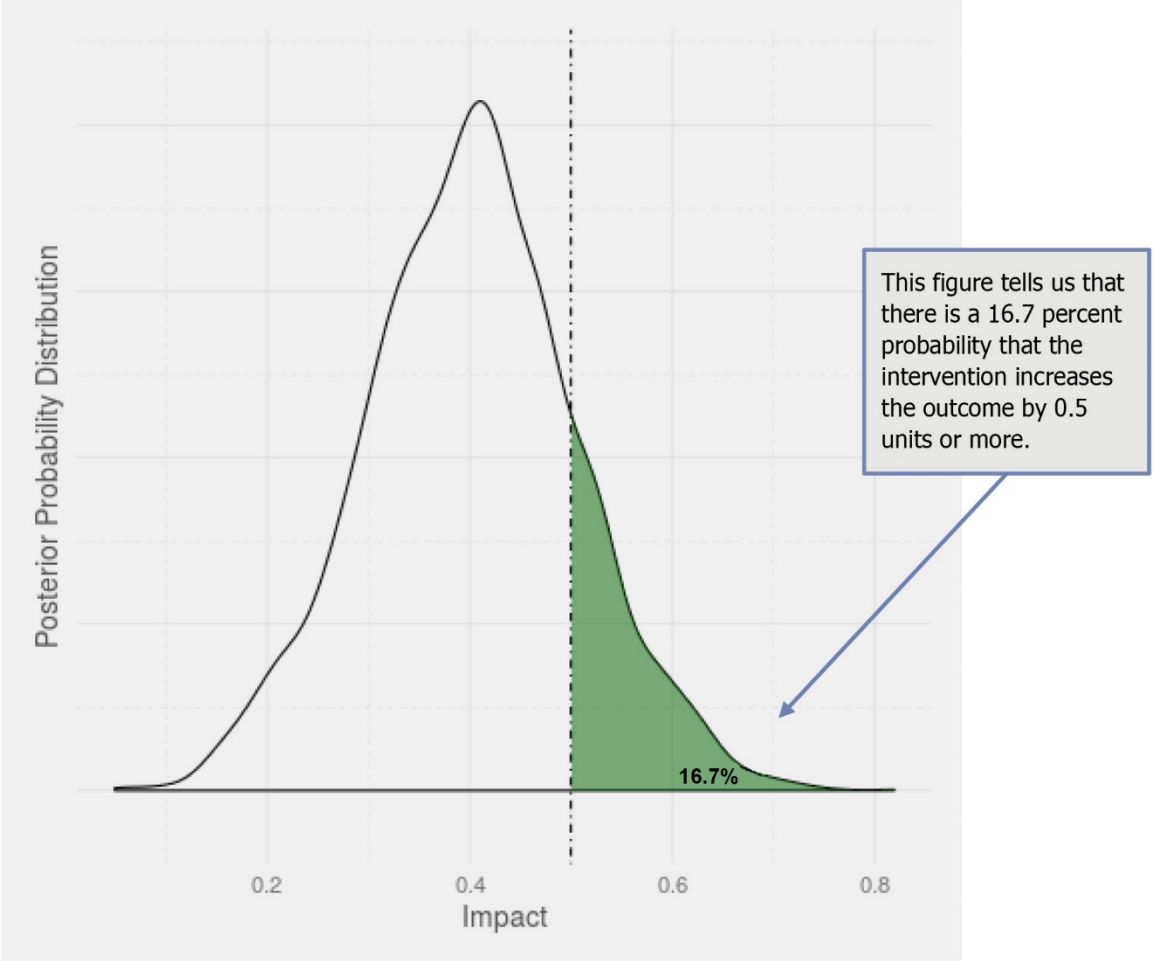
The Vocational Rehabilitation (VR) Program Evaluation Coach will help you determine if a VR service is having the desired outcome. Before you upload your data, the VR Program Evaluation Coach assumes that negative and positive effects are equally likely and that small impacts are more likely than large impacts. This is what statisticians call the prior distribution, as depicted in Exhibit 1. A prior distribution describes the analyst's beliefs about a measurable outcome before any data are taken into account.

Exhibit 1. Prior distribution of the effect of the service on the outcome



When the data are uploaded and the other inputs are specified, the VR Program Evaluation Coach calculates the probability that the effect of a service is above or below a threshold selected by the user (in Exhibit 2, the threshold, as indicated by the vertical line, is 0.5). This is the probability that the service is producing the desired outcome, as seen in Exhibit 2.

Exhibit 2. Probability that the service increases outcomes by 0.5 units or more, given the data uploaded



The Findings Brief for each program evaluation includes this information.

In this example, the user told the VR Program Evaluation Coach that any increase in the outcome greater than 0.5 units would be considered a success. The VR Coach calculated a 16.7 percent probability that the service had the desired impact by at least that amount.

It is important to note that the answer to whether the service is producing the desired impact depends on the threshold chosen by the user (0.5 units in this example) and how much uncertainty the user is willing to tolerate. For example, one user might conclude that a service is having the desired impact if there is at least a 75 percent probability that the service meets the selected threshold, whereas other users might want to be more confident, leading them to select a higher probability. In the previous example, if the user sets a certainty level of 95 percent, then the service would not be found to be having the desired impact because 95 is greater than 16.7.

VR Program Evaluation Coach

THE VR PROGRAM EVALUATION COACH

The VR Program Evaluation Coach uses a program called Stan (see Carpenter et al. 2016) to estimate a linear model by using Bayesian statistics. In particular, the VR Coach uses the R package RStan. The VR Coach acts as a simple graphic interface to help the user construct a formula. The user needs to select the outcome, treatment indicator, and any additional covariates that he or she might want to include. Before estimating the effect of the intervention, the VR Coach temporarily standardizes the data by using the R function "scale." Then, the VR Coach estimates the following model:

$$y_i = \alpha + \eta T + \beta x_i + \varepsilon_i$$

$$\varepsilon_i \sim N(0, \sigma)$$

$$\sigma \sim N(0,1)$$

$$\beta_k \sim N(0,1)$$

$$\eta \sim N(0,1)$$

where y_i is the outcome of interest for individual *i*, T is the treatment indicator, and x_i is a matrix with the other covariates.[1] The term η tells you the effect of the service—that is, how much greater the outcome was for the average individual receiving the service, $T = 1$, compared with the average individual not receiving it, $T = 0$. The VR Program Evaluation Coach computes the posterior distribution of η, which describes our understanding of the service's effect after observing the data. Using this distribution, the VR Coach calculates the probability that the treatment effect is above or below the threshold selected by the user, as shown in Exhibit 2.

WHICH VARIABLES SHOULD I INCLUDE AS CONTROL VARIABLES?

You should include all the variables, measured before you implement the service, that you think might affect the outcome of interest and for which you have good data. In the case of client employment outcomes, it's common to include previous employment experience and other characteristics, such as indicators for age, gender, disability category, socioeconomic status, and so on.

In addition, sometimes you might have to include an element in the model to account for clusters in the data, such as counselors or offices. This applies when individuals are assigned to receive or not receive the service as groups rather than as individuals. To estimate the effect of the intervention correctly, you will have to tell the VR Program Evaluation Coach that you used clusters during Step 4

[1] In the clustered case, we add cluster-specific random effects to the regression model that's given.

"Analyzing Your Data." By giving the VR Coach this information, you enable it to account for the possibility that individuals belonging to a cluster do better or worse just because they belong to that cluster and not because they are receiving the service. For example, when entire offices are assigned to conditions, the quality of counselors in each office can affect outcomes, along with any effect of the service. In this case, you would want to indicate the variable that identifies office clusters so that the VR Coach can account for office-based differences in outcomes.

The code for the VR Program Evaluation Coach is open-sourced under the General Public License Version 3 license and will be available soon on our GitHub repository.

CITATIONS

Carpenter, Bob, Andrew Gelman, Matt Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Michael A. Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. "Stan: A Probabilistic Programming Language." *Journal of Statistical Software*, in press.

Stan Development Team. "RStan: The R Interface to Stan, Version 2.10.1." Available at http://mc-stan.org.

Appendix B:
Glossary and FAQ

Glossary

Attrition One or more participants or groups of participants drop out of the study sample. Attrition is particularly important for randomized pilots because a high dropout rate can mean that the two groups that were equivalent or balanced when they were randomly assigned are no longer balanced. It is important to consider both overall attrition (the dropout rate of the full sample) and differential attrition (the difference in the dropout rate between the two groups).

Background Characteristics Information on study participants—such as gender, age, or type of disability—that might be related to receipt of the service or with the outcome you are measuring. Background characteristics are recorded as variables in your data set and are used to create comparison groups that are similar to each other.

Baseline The beginning of the program evaluation, before the introduction of the intervention (that is, just before the service is received).

Baseline Equivalence A demonstration of the similarity between the service recipients and nonrecipients before you implement the service. You achieve equivalence by comparing the average of each group on one or more observed characteristics.

Binary Variable See Dummy Variable.

Categorical Variable A variable that has a limited number of specific values. For example, race categories of white, black, and other may be represented as 1, 2, and 3, respectively.

Causality The relationship between a cause and an effect. The VR Program Evaluation Coach is designed to help you use program evaluation designs that enable you to assess whether the service you are studying leads to the outcomes you wish to achieve.

Cluster A group of individuals. For example, clients are clustered under counselors and within offices, and offices are clustered within regions. It is important to account for clusters in analyses when you are assigning service receipt based on groups, because individuals within a group share experiences. For instance, a set of clients receives services from the same counselor, and the quality of that counselor's support will affect clients' outcomes in a way that differs from that of a group of clients served by a different counselor.

Comparison Group A set of participants who are similar to the service recipients but who do not have access to the service. In a randomized controlled trial, the comparison group is the set of participants who were not selected to receive the service. In a matched comparison group design, the comparison group is usually the set of participants who do not receive the service.

Confounding Factor A hidden factor that influences an outcome (such as motivation or time), when you think the cause is something else (such as the service).

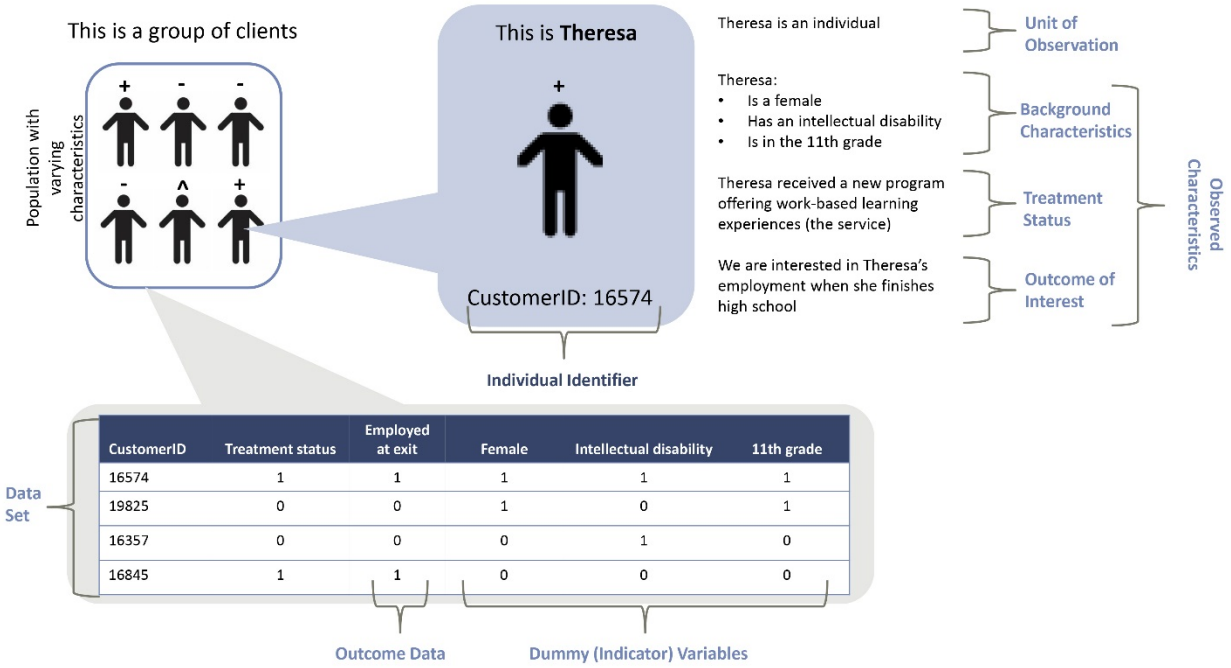
Control Variables See Covariates.

Covariates Factors or characteristics that differ across participants and may be related to the outcome of interest. It is important to account (or control) for these characteristics in your analysis of a service's effectiveness because one of these factors or characteristics—rather than the service—could explain your findings.

CSV File A CSV (comma-separated values) file stores data in tables as plain text and is widely used to exchange or transfer information. The VR Program Evaluation Coach uses the CSV format to read in your data files for analysis. You may save Microsoft Excel files as CSV files by using the "Save As" function.

Data Set A collection of related information that is organized in a table of rows and columns. Each column represents a variable (for example, age, type of disability, and whether a participant received a service). Each row represents a different observation (for example, a client, a counselor, or a provider). The VR Program Evaluation Coach uses your data set to run analyses and to determine whether a service is having the desired outcome.

Exhibit 1. Pieces of data



Dummy Variable Variable that takes the value of 0 or 1 to indicate the absence (0) or presence (1) of a specific trait or background characteristic. For example, to include gender in your data, your dummy variable would be female, with males assigned a 0 and females assigned a 1.

Effect The amount by which the average outcome of the treatment group differed from the average outcome of the comparison group. The difference is how much the service had the desired outcome.

Effectiveness The extent to which something succeeds in leading to the desired result or outcome. The effectiveness of a service depends on an agency's goal for that service. For example, one agency might determine that a service is effective if it reduces the amount of time between application and a signed IPE by at least three days, whereas another agency might decide that the service is effective if it reduces the amount of time between application and a signed IPE by any number of days.

Experiment A controlled test of the effect of an intervention (for example, a training in motivational interviewing) on participants' outcomes. A random assignment pilot is an experiment.

Impact See Effect.

Indicator Variable See Dummy Variable.

Individual Identifier A unique code, number, or ID assigned to individuals. These could be office IDs, counselor IDs, or client IDs. Unique values for each individual enable you to merge data easily and make information anonymous. For random assignment pilots, an individual identifier is also needed for randomly assigning recipients and nonrecipients.

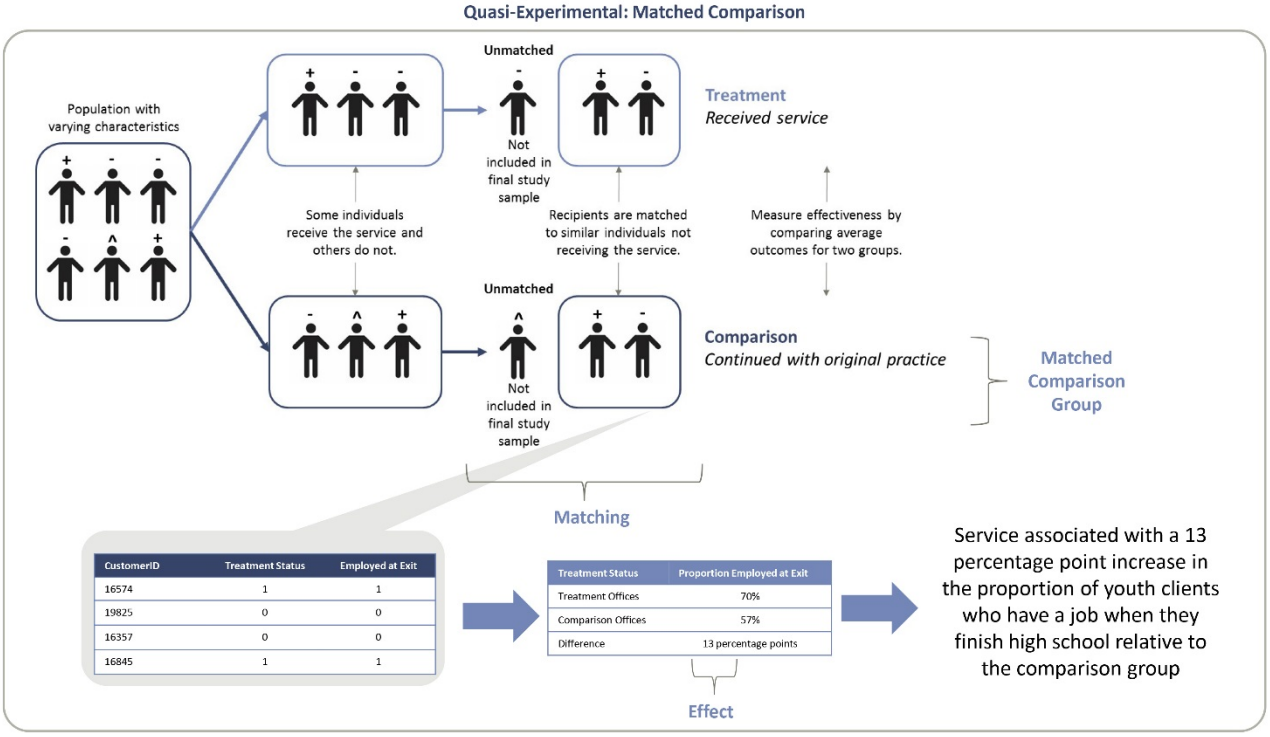
Intervention The service or program being tested. In a program evaluation, you investigate whether the intervention had an effect or an impact on participants' outcomes.

Intervention Group See Treatment Group.

Matched Comparison Group A set of participants, in a matched comparison experiment, who are similar to the service recipients but do not have access to the service.

Matched Comparison Group Design A program evaluation design that uses participants' pretest (if applicable) and background characteristics to create two sets of similar participants. One group receives the service (the treatment group) and the other group does not receive the service (the comparison group), and the outcomes of the two groups are compared. See Exhibit 2.

Exhibit 2. Pieces of a matched comparison group design



Matching The process of finding similar nonrecipients for comparison with service recipients. The analysis will compare only recipients and nonrecipients who are successfully matched. See the "Matching Overview", (https://vrevalcoach.com/static/pdf/Matching_Overview.pdf) guide for more information.

Missing Values A data point with no recorded information. For example, if you do not have an assessment score for a particular client, the absence of the score would be a missing value. In a spreadsheet, missing values may appear as blank cells or be tagged with labels such as NA, 999, or 0. You should make sure that all missing values are labeled with NA before uploading your data to the VR Program Evaluation Coach.

Numeric Type Data recorded as numbers.

Observation A single row in a data set. Each row usually represents a client, counselor, provider, or office.

Observed Characteristic A variable that you can measure and that can affect the outcome of interest. Examples of observed characteristics are age, type of disability, or prior work experience.

Outcome Knowledge, skills, attitudes, or other desired benefits attained as a result of an activity. An outcome could include, for example, employed at closure, receipt of pre-employment transition services, or secondary school referrals.

Outcome of Interest What you hope to change as a result of implementing the service, such as increased employment rates, higher earnings, or measurable skill gain.

Pilot A small-scale, short-term introduction of a service. In a pilot, only a portion of potential recipients receive the service. The purpose is to help you learn how the service would work on a larger scale before investing the time and resources in full implementation.

Probability The likelihood of something happening.

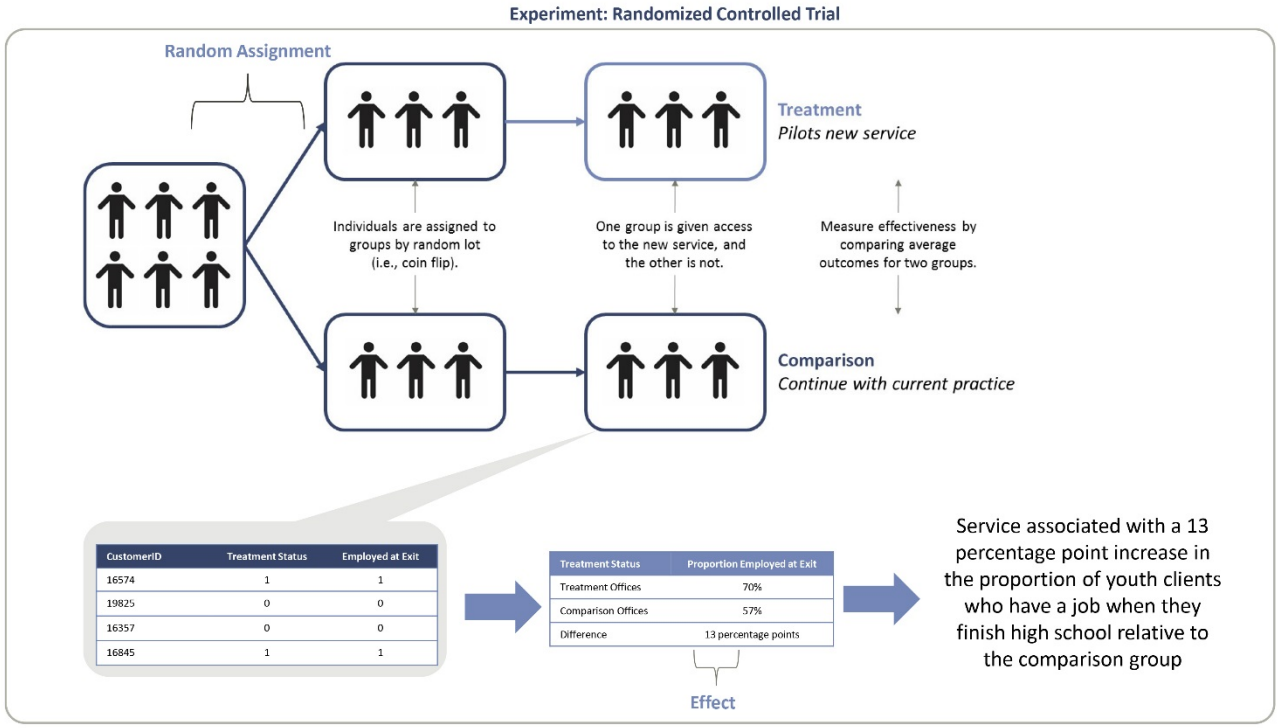
Program Evaluation Question An answerable question about the effect that a service could have on the recipients of the service. Defining a program evaluation question is the first step in a program evaluation.

Quasi-Experimental A design in which groups are created through a process that is not random. For a quasi-experimental design to be rigorous, the intervention and comparison groups must be similar, demonstrating equivalence on observed characteristics before the intervention begins. A matched comparison group design is a quasi-experimental design.

Random Assignment A process by which groups of recipients and nonrecipients are formed by chance (coin flip, random number generator, and so on). When carried out correctly, random assignment results in groups that, on average, are similar in both observed characteristics (such as age and gender) and unobserved characteristics (such as motivation) such that any differences in outcomes between the groups are attributable to the intervention alone. See the "Random Assignment Overview", (https://vrevalcoach.com/static/pdf/Random_Assignment_Overview.pdf) guide for more information.

Randomized Controlled Trial An experiment in which subjects are assigned at random to the group with access to the service or to the group without access to the service by, for example, a coin toss. See Exhibit 3.

Exhibit 3. Pieces of a randomized controlled trial design



Service Data Information concerning the actions and events that recipients engage in while receiving the service, or that VR staff or others engage in while implementing the service. Service data describe all types of recipient interactions with a service.

Statistically Significant The likelihood that a relationship between two or more variables is caused by something other than random chance.

Treatment Group Participants who are in the group with access to the service. In a randomized pilot, the treatment group is the set of participants randomly selected to pilot the service. In a matched comparison design, the treatment group is usually the set of participants who actually receive the service.

Treatment Status Whether a participant belongs to the treatment (service recipient) group or the comparison (nonrecipient) group. In a randomized controlled trial, the treatment status is the group to which the participant is randomly assigned. If a participant was assigned to the treatment group, the treatment status will always be treated, regardless of whether the participant receives the service or not. In a matched comparison group design, the treatment status usually refers to participants who received the service versus participants who did not receive the service. Treatment status is recorded in the data set as its own variable, with 1 indicating that the participant belongs to the treatment group and 0 indicating the participant belongs to the comparison group.

Unit of Observation The individual (could be clients, counselors, or providers), office, or region for which outcomes are examined.

Unobserved Characteristic Variable that you can't see or measure but that can affect the outcome of interest. For example, self-motivation is an unobserved characteristic.

Variable A characteristic, assignment, or outcome that can vary or change over time or across individuals. Variables are the columns in your data set. Examples of variables include treatment status, assessment scores, and demographic characteristics such as age and disability type.

Frequently Asked Questions from Vocational Rehabilitation (VR) Agencies

QUESTIONS ABOUT RAPID CYCLE EVALUATION

1. **What is a rapid cycle evaluation (RCE)?**

   RCEs use a rigorous, quick-turnaround approach to determine whether a service or implementation approach meets the needs of a VR agency or clients. The VR Program Evaluation Coach can help you conduct an RCE.

2. **Why is RCE important?**

   Conducting RCEs empowers VR staff to evaluate the effectiveness of changes to policies, programs, or services, allowing staff to make informed, evidence-based decisions in action-oriented time frames. With the Workforce Innovation and Opportunity Act (WIOA), the need for evidence-based decision making has increased significantly. Conducting RCEs by using the VR Program Evaluation Coach is one way to gather needed evidence.

QUESTIONS ABOUT GETTING STARTED

3. **Who should use the VR Program Evaluation Coach?**

   The VR Program Evaluation Coach is designed for VR staff who wish to evaluate a service or service change. The VR Program Evaluation Coach can test the effectiveness of a service that an agency already uses or help set up a program evaluation of a new service as part of a pilot process.

   Although the tool was developed for evaluating services, you can use the evaluation process to evaluate programs, interventions, and strategies.

4. **How do I begin using the VR Program Evaluation Coach?**

   Visit http://www.VREvalCoach.com to answer a few brief questions to create an account. Review all resources and follow the step-by-step process for setting up and conducting a program evaluation.

5. **What if I'm not ready to start a program evaluation right now?**

   You can access the tools and downloadable guides to learn more about the program evaluation process. You'll find all of the tools on the "Preview Tools" page. If you want a better understanding of program evaluation and what you need in order to conduct your own evaluation, the following guides are particularly helpful:

   - "VR Program Evaluation Coach Overview", (https://vrevalcoach.com/static/pdf/VR_Program_Evaluation_Coach_Overview.pdf)

- "What You'll Need to Get Started", (https://vrevalcoach.com/static/pdf/What_You_Need_to_Get_Started.pdf)

6. What information does the VR Program Evaluation Coach provide?

The VR Program Evaluation Coach tells you whether a change in VR policies, programs, or services had the desired outcome. The VR Program Evaluation Coach provides summary statistics for the data you upload for your analysis and the results from the analysis to determine whether the service is having the desired outcome. It also captures important contextual information about your agency, office, or region and how the service was implemented, so that you can easily share and interpret the findings. All information generated by the VR Program Evaluation Coach is based on data you input into the tool.

7. What do I need to use the VR Program Evaluation Coach?

The key ingredients for using the VR Program Evaluation Coach include:

- The service or service change you wish to test
- A group of service recipients and nonrecipients, or potential recipients and potential nonrecipients
- An outcome related to the service that you can measure, such as exiting from service receipt with employment or time to Individualized Plan for Employment
- A data set to be uploaded to the VR Program Evaluation Coach
  - If you are setting up a forward-looking pilot, you have to identify the service and data sources, but you may compile the information over the course of the pilot. You will start with a list of participants who will be assigned to pilot the service or not pilot the service.
  - If you have already implemented the service and want to undertake a backward-looking evaluation, you might have all of the needed information already in hand. You will need a list of service recipients and nonrecipients, an outcome measure for each recipient, and recipients' background characteristics.

8. How long should a program evaluation take to complete?

A typical time frame from start to finish for a program evaluation varies depending on the nature of the program evaluation, your program evaluation question, and the outcome of interest. It's possible for some evaluations to be completed in weeks, whereas others might need more than a year. If the outcome of interest takes more than a few months to affect or measure (such as exiting with employment), that will drive the timeline. Program evaluations are also meant to be iterative, so you could go through the evaluation process several times in a year.

9. What if I care about longer-term outcomes, such as exiting with employment, but want to see if things are working now?

Program evaluation and RCE can still be useful if intermediate outcomes are needed steps toward achieving the longer-term outcome. In the case of exiting with employment, it is possible that increasing job placement services or training services in the short term is a useful step toward

improving the rate of exiting with employment. You could use the VR Program Evaluation Coach to assess whether your service change is having an impact on the intermediate measures.

10. How much time will I need to invest?

The amount of time needed to conduct a successful program evaluation will vary with where you are in the evaluation process (for example, whether you have already chosen the service to be evaluated); the evaluation method; the number of stakeholders with whom you need to work; and how quickly you can collect and organize your data. If your data are already available, you could complete a program evaluation in as little as a few hours.

11. Whom should I involve in my program evaluation?

Consider including key stakeholders or decision makers and people with the skills necessary to help you complete your program evaluation. Our "What You'll Need to Get Started", (https://vrevalcoach.com/static/pdf/What_You_Need_to_Get_Started.pdf) guide will help you understand what you will need to complete your program evaluation. Ideally, your program evaluation team will include someone with experience in handling data sets. Often, people with these skills work in accountability or research offices. However, if you do not have access to experienced data analysts, the VR Program Evaluation Coach provides instructions for preparing your data. (No statistics experience is necessary.) Other desirable team members include an administrative champion and staff involved in service design or provision.

12. How much data analysis or statistics do I need to know?

With the VR Program Evaluation Coach, your statistical knowledge will not be a barrier to completing a successful program evaluation. The VR Program Evaluation Coach is designed for people with varying data analysis and statistical backgrounds. It will be helpful to have access to someone able to put together data sets, but user-friendly guides provide all the assistance you need to set up your data correctly. The VR Program Evaluation Coach's dashboard(s) will complete the statistical portion of your analysis for you. If you have no background in data analysis or statistics, you can refer to the guides in the VR Program Evaluation Coach to help you understand the findings produced for your program evaluation.

13. Can I compare two different services, or different implementation of the same service?

Yes. Sometimes it is not feasible, practical, or desirable to compare service recipients with nonrecipients of the same service. You might be more interested in learning about the best way to implement a service, or the best choice among two different services, and a program evaluation or RCE is an excellent way to do that.

You can compare two different services. For example, you might wish to compare two job placement techniques to see which produces a greater impact on exiting with employment. If you prefer one technique over the other, your program evaluation can designate the preferred technique as the treatment group's service. (For example, you might prefer one technique because it is less expensive, or anecdotal evidence might show that it appears to be effective.) The

VR Program Evaluation Coach can help you test whether one service is more effective than the other.

You can also compare two groups that use the same service in different ways. For example, one group of clients could receive one-on-one financial or benefits counseling with a counselor, and the other group could receive financial or benefits counseling in small groups. The VR Program Evaluation Coach can help you test whether one implementation strategy is more effective than the other.

In both cases, you will be able to tell only whether one approach was more effective than the other, not whether the service itself was effective. In addition, if you are comparing similar services or strategies, the two approaches could be equally effective and your results might be inconclusive.

14. What if I'm targeting use of a service to a specific set of clients? Can I still find out if the service is effective?

To assess the effectiveness of a service, it's critical to have a good measure of how recipients would have performed without it. This can be challenging when the service is targeted to a specific set of clients. For example, if you offer a specific program to all youth clients with developmental disabilities, you would not have a comparison group and therefore would not be able to draw strong conclusions about the service's effectiveness. However, if you have a capacity constraint—you don't have enough counselors to offer the program to all youth clients with developmental disabilities or you have to introduce the service over time across the state—you could set up a randomized pilot that would enable you to assess effectiveness. Even when you use a service with a specific set of clients, you could use the VR Program Evaluation Coach to evaluate different implementations of the service or different strategies for incorporating the service for those clients. For example, you could randomly select some of the youth clients with developmental disabilities to receive a more intensive version of the service.

QUESTIONS ABOUT PROGRAM EVALUATION DESIGN

15. What is the benefit of random assignment over the other designs?

The goal of a program evaluation is to determine the impact of the service based on differences in outcomes between recipients and nonrecipients. To be confident that those differences are solely attributable to receipt of the service, you will have to reduce the chance that something else is affecting the outcome. Random assignment is the best way to reduce the chance that anything but the service affects the outcome; the reason is that recipients and nonrecipients are similar on both observed characteristics (such as age and gender) and unobserved characteristics (such as motivation). Matching, the other design supported by the VR Program Evaluation Coach, enables you to create groups that are similar only on observed characteristics. Therefore, the benefit of random assignment is that it gives you the highest level of confidence in your findings.

16. Is random assignment unfair?

You are likely implementing a service because you assume that it will benefit recipients. With some services, however, you might not know for sure if they will succeed in your context and with your clients. The best way to learn whether the service works as intended is to pilot it—or offer it to a small group of recipients on a trial basis. In any pilot, some people will receive the service first, while others will have to wait to receive it. A fair way to determine who receives and does not receive the service is to use a random chance process, (https://vrevalcoach.com/static/pdf/Random_Assignment_Overview.pdf) (such as flipping a coin). If the service has an impact, more people can get access to it when you know that it will benefit them.

17. What questions can I answer by using the VR Program Evaluation Coach?

The VR Program Evaluation Coach will enable you to answer the following question: Is the service having the desired impact? The VR Program Evaluation Coach can answer this question based on any type of recipient and any outcome of interest you are measuring.

18. How large does my sample have to be? My office might be too small for the VR Program Evaluation Coach.

Many factors contribute to the recommended sample size, including whether you are interested in individuals (such as clients) or groups of individuals (such as all clients served by one counselor); how large an effect you expect the service to produce; and how much the characteristics of clients, VR staff, or providers vary. In general, a larger sample is preferable. The more clients, VR staff, or providers in your sample, the more likely that you will receive an answer to your program evaluation question (rather than inconclusive results). However, samples do not necessarily have to be huge, and some data are better than no data.

19. I am introducing a service to all clients with autism. What type of program evaluation can I undertake?

To conduct a program evaluation, you must be able to compare a group receiving the service to a similar group not receiving the service. If all clients with autism are receiving the service, you would not have any nonrecipients with whom to compare them. You could consider introducing the service in phases; that way, you would have a short period during which you have recipients and nonrecipients. You could also implement the service in different ways and compare outcomes to support one implementation method over the other. For example, some clients could receive the service once a month; others could receive it weekly. Depending on how you determine who receives the service for a specific period, you could use the randomized pilot or matched comparison pathways offered by the VR Program Evaluation Coach.

20. If all of my clients are using the service, can I use service delivery data to categorize my treatment and comparison groups—by designating low-usage users as my comparison group?

Though such an approach gives you some information, it does not provide the convincing evidence you need to conclude that the service is having the desired outcome. This approach, however,

could be a good starting point for developing further ideas to be tested in a subsequent pilot. The reason is that assigning clients to groups based on their service receipt increases the likelihood that unobserved differences in clients will affect outcomes. Other factors could lead some clients to receive less of the service, and those factors might also affect outcomes. For example, lack of interest or motivation could lead to a client's lower use of a service. Those factors could also be associated with poorer outcomes. If that is the case, you wouldn't be able to determine whether it was low usage or motivation that caused the observed outcomes.

21. Does the VR Program Evaluation Coach enable me to analyze the effects of the service on subgroups, such as transition-age youth or individuals with specific disabilities?

You can analyze subgroups by using the VR Program Evaluation Coach; however, to do so with the current version of the VR Program Evaluation Coach, you will have to perform each step of the VR Program Evaluation Coach for each subgroup of interest. You will need a data set specific to each subgroup in order to analyze outcomes by subgroup. (For example, to analyze effects by gender, you will need one data set with only male clients and another data set with only female clients.)

22. Can I look at several outcomes?

The VR Program Evaluation Coach is currently designed to analyze one outcome at a time. However, you can easily select one outcome, obtain results, produce a findings brief, and then repeat that process for a second outcome.

23. One counselor (or one office in my region) is implementing a service. Can I compare the clients served by the counselor (or by the office in my region) with clients served by other counselors (or by other offices)?

To conduct a program evaluation, you need to compare a group receiving the service with a similar group not receiving the service. If only one counselor or office is using the service, you will not be able to disentangle the effect of the service from the effect of characteristics that are unique to that group (such as the quality of the counselor or office). However, if you have data from a situation like the one described here, you can use the information to start building evidence. Comparing results from one counselor or office with another group can provide suggestive evidence. To build more confidence in the results, we recommend that you set up a forward-looking pilot that includes more than one counselor or office in each group.

24. Can I compare clients who use the service this year to clients who did not use the service last year?

The VR Program Evaluation Coach does not currently support such comparisons for **causal** analysis, but you can use the software for an evaluation that uses an other comparison group/matched comparison design that provides correlational evidence. Environments change over time. If you compare clients in one year to clients in the following year, any differences could be attributable to differences in the clients or to other programs, policies, services, or contextual factors that changed from one year to the next. However, if you have data from more than one year, you can use the information to start building evidence. Comparing results from across years

can provide suggestive evidence. To build more confidence in the results, we recommend that you set up a forward-looking pilot that compares clients from the same year.

QUESTIONS ABOUT INFORMATION SHARING

25. How can I share the results of my evaluations?

The VR Program Evaluation Coach enables you to generate a portable document format (PDF) file of your findings or print the results from any web browser. We recommend that you follow existing protocols in your office or agency for sharing results.

26. How does the VR Program Evaluation Coach handle client information?

The VR Program Evaluation Coach does not require personally identifiable information (PII) in the data set that you upload (called user data files in the Privacy Policy) in order to use the tool. The VR Program Evaluation Coach provides guidelines on how to prepare and upload data sets (user data files), including specific measures for de-identifying clients and avoiding PII. However, the VR Program Evaluation Coach cannot prevent you from uploading data sets that include PII. As you prepare to create your program evaluation report or brief, the VR Program Evaluation Coach will ask you to enter information (called user input in the Privacy Policy). The VR Program Evaluation Coach will provide tips along the way to help you avoid inadvertently entering PII. Finally, the program evaluation reports or briefs produced by the tool (called user output in the Privacy Policy) will contain only information that is sufficiently aggregated so that it is nonidentifiable.

27. What happens to the data files that I upload into the VR Program Evaluation Coach?

The data sets or data files you upload (called user data files in the Privacy Policy) are stored exclusively in a temporary cache. They are not stored permanently anywhere in the tool. The data contained in your user data files will be used only to produce the analysis, brief, or report that you request. The user data files are not used for any other purpose. When you are finished and leave that page, your user data files are permanently erased from the temporary cache within 15 minutes.

28. What happens to reports after completion of the program evaluation? Who owns the information?

You always own all of the information that you put into the VR Program Evaluation Coach and all of the information that comes out of your use of the VR Program Evaluation Coach. However, the Terms of Use and Privacy Policy require you to give Mathematica Policy Research the right to use of the information, so long as it is aggregated and/or de-identified. The information you enter into the VR Program Evaluation Coach (the Privacy Policy calls it your user input) and the evaluation reports and briefs created by VR Program Evaluation Coach (the Privacy Policy calls this your user output) can be used for educational, academic, and research purposes and to help other users or to promote research and development of technology like the VR Program Evaluation Coach. Your user input (information you entered) and user output (your evaluation reports and briefs) are

stored permanently on Amazon Web Services under your VR Program Evaluation Coach account. Remember, the data set or data files that you upload (the Privacy Policy calls these your user data files) are not stored anywhere, but they are deleted when you finish using the VR Program Evaluation Coach. Please visit our Privacy Policy, (https://vrevalcoach.com/PrivacyPolicy) and Terms of Use, (https://vrevalcoach.com/TermsOfUse) for more details.

QUESTIONS ABOUT GETTING ADDITIONAL ASSISTANCE

29. I want to know what I am getting into. Is there an overview of all of the steps?

Check out "What You'll Need to Get Started", (https://vrevalcoach.com/static/pdf/What_You_Need_to_Get_Started.pdf) for an overview of the skills and data you will need to conduct a successful program evaluation. You can also find a list of all of the tools under "Preview Tools", (https://vrevalcoach.com/PreviewTools)

30. A lot of terms are unfamiliar to me. Is there a printable glossary that I can use as a reference?

Yes. The VR Program Evaluation Coach uses technical terms only when absolutely necessary. In these cases, there is often a definition that you can view by placing your cursor over the word. For more extensive definitions, you can view the printable glossary, (https://vrevalcoach.com/static/pdf/Glossary.pdf). The technical vocabulary is commonly used in program evaluation and statistical research. Familiarity with these terms will help you understand other research as well.

31. I don't have the answers to all of the questions.

a. Are there some responses that are required and others that are not?

Yes. Answering all of the questions will enable you to produce the most complete findings brief, but only some questions are required for completion of the analysis. Some tools will be locked (and allow you just to peek at the contents) until you answer required questions. However, if you skip many questions, it will likely be difficult to interpret and share your findings.

b. Can I skip a question and return later?

Yes. The VR Program Evaluation Coach is designed so that you can move through and answer questions when you have answers.

c. Will the VR Program Evaluation Coach remember which questions still have to be completed?

Yes. The VR Program Evaluation Coach tracks the questions you have completed and prompts you for key information when it is required.

32. Whom do I contact if I need help using the tool?

You may leave questions and your contact information in the Feedback tab on the right margin of each page. Someone from the VR Program Evaluation Coach project team will then contact you. You may also periodically check the frequently asked questions for updated responses.

VR Program Evaluation Coach

33. How can I pilot the VR Program Evaluation Coach and receive technical assistance?

The VR Program Evaluation Coach is freely available for anyone to use. Resources are available to support a limited number of agencies in conducting program evaluations with the VR Program Evaluation Coach. Contact the VR Program Evaluation Coach evaluation team through the email on the website if you are interested in such support.