# Tool 1.4 – Sustainable Data Use

Strengthening Analytics in Government Agencies: A Toolkit for Sustainable Data Use offers strategies and tools for individuals in government agencies and other organizations with similar needs.[1] The material was sourced from interviews with practitioners who have successfully built sustainable data use into their everyday practices. It covers a variety of subjects—from staffing and technology to collaboration and funding—that can impact the longevity of analytics work in the public sector. While the toolkit was developed with state TANF agencies in mind, many of the techniques offered may also be useful for individuals in a range of government agencies and other organizations, where the challenges—and potential for impact—are similar.

1. Wiegand et al. (2023).

# Phase 2

# Accessing the Data

The second phase of data analytics work includes activities to address legal, ethical, and cross-agency coordination considerations to ensure data access. The TDC Pilot Initiative expected all pilot agency teams to already have (or be close to having) access to administrative data from the TANF program and data about earnings and employment outcomes from the unemployment insurance system (UI wage data) from the beginning of the pilot. Therefore, pilot agency teams were able to spend more time preparing and analyzing the data than accessing the data. Although most pilot agency teams had access to much of the needed data, the TDI team developed Expanding TANF Program Insights: A Toolkit for State and Local Agencies on How to Access, Link, and Analyze Unemployment Insurance Wage Data.[1] This toolkit was created to help TANF professionals develop more robust, data-driven practices using administrative data on earnings for program improvement purposes. It may also be useful to other state human services agencies (for example, Child Support) that want to expand their data use, as well as policymakers interested in supporting improved workforce outcomes. State Department of Labor agencies may also gain useful insights from the data preparation section, as well as from the broader discussion of ways to use employment data to improve human services programs.

The toolkit's primary purpose is to offer practical guidance on how to access, link, and analyze employment data from unemployment insurance systems for program monitoring, reporting, and evaluation. Further guidance can be found in the companion GitHub repository, which offers open source code and documentation for staff preparing employment data for analysis.[2] It also includes resources with related supplemental materials that have emerged from this project. Resources cover a variety of topics, including equity, data security, and programming and data quality control tips.

1. Yang et al. (2022).
2. Yang et al. (2022).

# Phase 3

# Preparing the Data

"Garbage in, garbage out," as the saying goes: Flawed input data produces flawed outputs. Carefully preparing data for analysis is an important step toward ensuring that it meets certain quality standards. Preparing data for analysis in this phase can include various procedures related to linking, de-identifying and restructuring the data so that it is ready to be analyzed. See the UI Wage Toolkit for guidance on those procedures and the Applied Data Analytics training that provides resources and instruction on activities related to preparing and analyzing the data.[1] The tools in the current toolkit focus on the quality checking and cleaning activities.

**Tool 3.1 – Data Quality Control Checklist** is useful for documenting features of the data, including any limitations of the data for analysis. It can help the team reach a shared understanding of the data and can help future staff get up to speed quickly in preparation for working with the data. Tool 3.1 organizes questions related to data quality checking for staff to consider in five categories: (1) the nature and source of the data, (2) prior to processing, (3) prior to writing code, (4) while processing, and (5) after processing.

**Tool 3.2 – Template for UI Wage Data Quality Control Memo** can be used to document information about a file including UI data and checking for quality. Topics include: (1) project background, (2) file locations, (3) key decisions, (4) checking a raw data file, (5) checking data updates, and (6) checking a person-level file. Although this template is based on using UI wage data, many of these topics are relevant for checking the quality of any kind of data.

1. Coleridge Initiative (n.d).

# Tool 3.1 – Data Quality Control Checklist

To ensure that data meets certain quality standards before analysis begins, this checklist offers questions to consider when checking data for quality. Once each question is answered or considered, you can check the box on the right.

**Depending on the source of your data, some questions you may want to ask the data provider:**

- Where do the data derive from? ☐
- What are the data used for? ☐
- How will the data be extracted? Is there a standard process for this, or will this need to be developed for this purpose? ☐
- What fields will the data include? Is there any documentation available? ☐
- Is there a specific set of records that will be selected for extraction? How will that be specified? ☐
- Is there a lag in the data entry that would require a wait to get complete data through a particular period? How often and when are the data updated? ☐
- Are there any system conversions that took place during the period these data cover? ☐
- How far back do the data go? ☐
- What are the known weaknesses in the data? ☐

**Prior to processing the data:**

- What are the criteria for inclusion in the dataset? ☐
- What is the sample size? ☐

**Prior to writing code to process the data:**

- What data checks should be made to ensure accuracy of the final results? ☐

**While processing the data:**

- Did I check for missing data? Also, did I check by subgroups, to uncover patterns of missing data? ☐
- Did I check for data conversion issues? ☐
- Did I check for duplicate or partial duplicates? ☐
- Did I check for internal inconsistencies, unexpected values, and outliers? ☐
- Did I check for common programming mistakes while:
  - reading data files? ☐
  - merging files? ☐
  - restructuring data? ☐
  - creating variables and date values? ☐
  - working with arrays? ☐
- Did I select the correct sample(s) for routines that apply to a subset of records? ☐
- Did I apply the correct formats (numeric, character, date) for all variables? ☐
- Did I check sample size for each outcome measure? ☐
- Did I create summary/aggregate checks? ☐
- Did I follow a group of random cases through every programming step? ☐
- Did I check that these data files look consistent over time? ☐
- Do categorical measures add up to 100%? ☐
- Did I follow up on any suspicious findings? ☐

**After processing the data:**

- If datasets are updated over time, did I review all data checking output with each data update? ☐
- Did I clearly document any data quality issues found and resolved? ☐

# Tool 3.2 – Template for UI Wage Data Quality Control Memo

This template for a data quality control (QC) memo outlines key information that should be documented for a data analytics project that uses Unemployment Insurance (UI) wage data collected from a state or federal government source. The purposes of this type of QC memo are: (1) to document the quality of the data files for future reference, (2) to share information in an easy-to-digest format with other individuals you work with, and (3) to highlight potential issues you may want to ask the data provider about. There are several notes throughout the template and placeholders where information can be filled in.

The template assumes that those who use it already have a baseline knowledge of UI wage data. More information on UI wage data can be found in Expanding TANF Program Insights: A Toolkit for State and Local Agencies on How to Access, Link, and Analyze Unemployment Insurance Wage Data.[1]

Although this template is based on using UI wage data, many of these steps are relevant for checking the quality of any kind of data. Depending on your programming capacity, this kind of QC memo can often be autogenerated from statistical software, with the software producing the numbers presented in brackets.

1. Yang et al. (2022).

**[Programmer Name/Author]**

**[Date]**

**Project Background**

[Project Name] is [description of project and purpose of using UI wage data].

This memo discusses the contents and quality of the UI wage data file received on [Date file received] that covers [First Quarter/Year on file] through [Last Quarter/Year on file]. This is the [number of files] received for the project.

**File Locations**

The data processing programs and datasets are located at the following paths shown in Table 3.2.1:

## Table 3.2.1 File Locations

| STAGE | TYPE | FILE PATH | PROGRAMMER |
|---|---|---|---|
| Checking Returned Data File | Program | | |
| | Dataset | | |
| Checking Data Updates | Program | | |
| | Dataset | | |
| Checking Person-Level File | Program | | |
| | Dataset | | |

**Key Decisions**

*Note: Below are examples of key data quality issues and decisions that teams may need to make depending on the quality of the data file. You can include decisions like these or any other decisions that you make here. This section provides a high-level summary of the quality of the data and the steps taken to address any issues. The decisions documented here will depend on the identified problems. Some examples are outlined below.*

- [The file contained X number of exact duplicates (duplicates on all fields). For these duplicates, one of the records was dropped as they were thought to be a mistake.]
- [The file contained Y number of partial duplicates defined by having the same SSN, quarter, and earnings amount, but a different employer ID. For these duplicates, one record was dropped when it was thought the employer ID had changed over time.]
- [The file contained Z number of partial duplicates defined by having the same SSN, quarter, and employer ID, but a different earnings amount. For these duplicates, the mean

of the two earnings amounts was taken and filled in on one of the records. The other record was dropped.]

- [The (number of outliers) records had earnings amounts greater than (threshold for outlier checks). At this time, no changes were made to these records. They were flagged so that the analysis could be run with and without them included, as a sensitivity check.]

**Checking Raw Data File**

The request file sent to the UI agency included [number of SSNs sent] SSNs. The returned file includes fields for [SSN, Quarter/Year, Employer ID, NAICS code, and earnings amount]. Data were returned at the [person-employer-quarter] level.

- **Record counts:**
  - There are [number of records] records and [number of unique SSNs] unique SSNs in the returned data file. [100*(number of missing SSNs)/(number of SSNs sent)] percent of the SSNs on the request file do not have a matched record in the returned file. [Insert text on whether this is expected, or unusually high/low.]
  - The earnings on this file total [sum of earnings on entire file].
- **Missing or invalid data:**
  - There are [number of missing earnings amounts] records with missing earnings amounts, and [number of invalid earnings amounts] records with negative or zero earnings amounts in the returned file. [If there are missing or invalid data, insert text on why and whether/what to do about it.]

    *Note: In UI data, missing earnings amounts often simply mean that a person is not working in that quarter. Missing amounts do not necessarily indicate a problem if the incidence is within reasonable expectations given one's knowledge of the target population.*
  - There are [number of missing quarters] records with missing quarters in the data. The missing quarters are [list of missing year/quarter variables].
  - There are [number of invalid quarters] records with quarters that fall outside of the date range expected on the returned file. These quarters account for [number or records with invalid quarters] records in the file. [Insert text on what to do about missing or invalid quarters.]
- **Duplicate records:**
  - There are [number of exact duplicates] exact duplicates in the file. [If there are exact duplicates, insert text on what was done to address them.]

*Note: Exact duplicates are records that contain the same values for every field. For example, the same SSN, quarter, employer ID, and earnings amount.*

*Note: If the file has more than a few exact duplicates, you may want to check with the data provider for the source of these exact duplicates.*

- There are [number of partial duplicates] partial duplicates by [fields used to check for partial duplicates] in the file.

  *Note: Partial duplicates occur when there is more than one record that has the same value for some (but not all) of the fields. There may be different variations of partial duplicates, depending on what fields you have in the returned file. Look at each variation and insert text under this bullet about what you did for each type of partial duplicate. Examples of partial duplicates are:*

  *- Same SSN, Quarter, NAICS code, Employer ID, different earnings amount*

  *- Same SSN, Quarter, NAICS code, earnings, different Employer ID*

  *- Same SSN, Quarter, Employer ID, earnings amount, different NAICS code*

  *Note: Partial duplicates may appear when you receive multiple files from a UI agency that cover overlapping quarters (described more below).*

- **Outliers:**

  - Per quarterly record: There are [number of earnings amounts ≥ $20,000] records with earnings amounts of $20,000 or more ([100*(number of high outliers)/(total number of records)] percent of all earnings records in the file). [Insert text about how you handled quarterly outliers.]

    *Note: The definition of a potential high earnings outlier will vary depending on who is in your sample. You should determine what you would consider an outlier based on expectations for the study population.*

  - There are [number of earnings amounts ≤ $20] earnings amounts of less than $20 in the file ([100*(number of high outliers)/(total number of records)] percent of all earnings records in the file). [Insert text about how you handled quarterly outliers.]

    *Note: Teams often decide not to make any corrections to low earnings outliers, as they typically will not make a difference in the analysis and are considered plausible earnings amounts.*

  - Per person: There are [number of summed earnings amounts across person/quarter ≥ $20,000] of individuals with more than $20,000 in total quarterly earnings, accounting for [100*(number of people with high outliers/total number of people on the returned file] of the individuals in the file. [Insert text about what to do about summed outliers by person/quarter.]

- **Consistency of record counts across quarters and SSN:**
  - The average number of individuals per quarter in all quarters is [average of counts of number of unique SSN over all quarters in the file]. [List of quarters with less than 95 percent of the average] have fewer than 95 percent of the average number of individuals per quarter. [List of quarters with more than 105 percent of the average] have more than 105 percent of the average number of individuals per quarter. [Insert text about whether these variations are expected or need more investigation.]
  - A more granular check of variation in earnings for the same individuals over time was also conducted. Earnings of individuals grouped by the first three digits of their SSN are [consistent or inconsistent, depending on variance of counts across quarters]. [Insert text about whether these variations are expected or need more investigation.]
- **Final record counts of cleaned returned data file:** After deleting invalid and duplicate records and handling outliers, the final returned data file has [number of records] earnings records that account for [total sum of earnings amounts] in earnings from [Start Year/Quarter] through [End Year/Quarter].

### Checking Data Updates

*Note: Teams often request and receive multiple shipments of UI wage data from UI agencies. This is often due to the availability of data (for example, some providers only maintain a certain number of quarters of data at a time). If possible, when requesting multiple files, it is good to request the same quarters of data on multiple files, as employers sometimes provide missing or updated records (earnings amounts or employer IDs) that will only be reflected in the latest file. For example, one file could cover Q1, 2000 to Q4, 2004 and a second file could cover Q1, 2002 to Q4, 2006. This section provides guidance on how to compare records from these overlapping time periods on a historical and current data file, as well as how to handle duplicates.*

- **Partial duplicates:**

  *Note: Count partial duplicates that are identical on every field except the file date. This is a check to make sure that earnings have not changed across files, so there should be a lot of these. For these sets of partial duplicates, remove record from the previous dataset.*

  *Note: Count partial duplicates that are identical on SSN and quarter but do not match on another field. The most common of these will be partial duplicates with different earnings amounts or different employer IDs. For the former, you may decide to keep the higher amount, the more recent amount, or take the mean of the two amounts. For the latter, you may decide to keep both records or only the more recent one.*

- **Differences in earnings amounts:**

  *Note: You will want to document differences in earnings amounts between the historical and current file. It is possible that employers corrected data they had previously submitted to the UI agency.*

- **Record count:**

  *Note: You should count the number of records after resolving duplicates and calculate the total sum of earnings in the merged, updated file. Then check that these amounts are consistent with what you expect.*

**Checking Person-Level File**

- **Record counts:**

  - There are [number of records] records and [number of unique SSNs] individuals in the person-level file.

    *Note: the number of records should match the number of individuals in the person-level file. These numbers should also match the number of valid SSNs that were sent to the UI agency in the request file.*

  - The earnings on this file totals [sum of earnings on entire file].

    *Note: the sum of earnings in this file should match the sum of earnings in the updated data file above.*

- **Counts of outlier earnings:** About [100*(number of individuals with earnings outliers)/(number of total individuals on person-level file)] percent of all individuals have outlier earnings in the person-level file.

  *Note: Table 3.2.2 shows an example of thresholds you could use to check for high earnings amounts. The threshold you use will depend on your study population.*

**Table 3.2.2 Counts of Outlier Earnings**

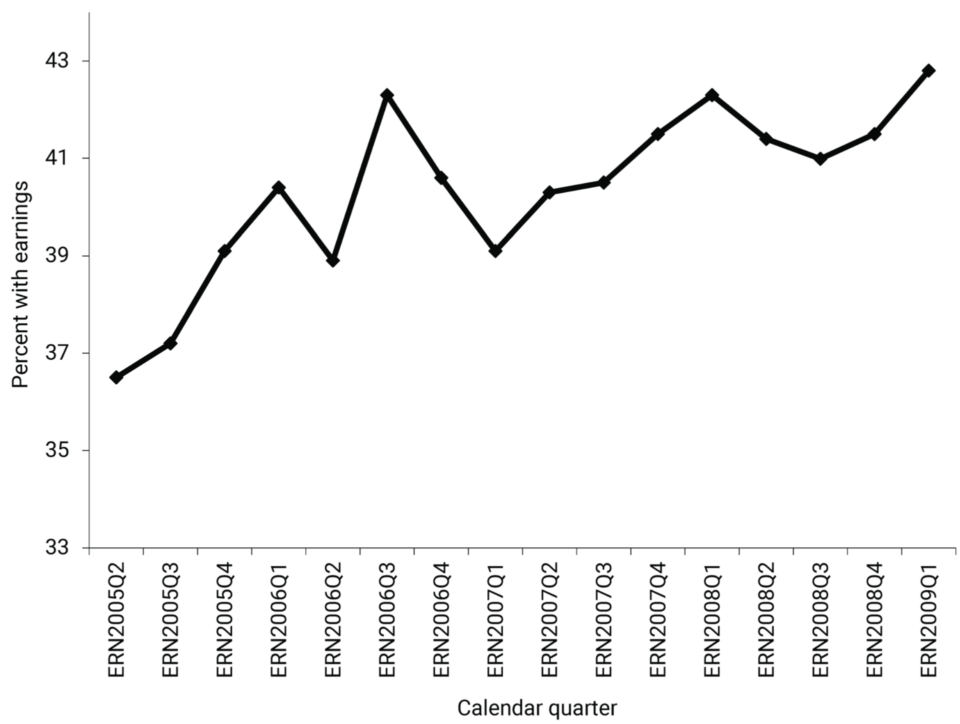
| EARNINGS AMOUNT (EQUAL TO OR GREATER THAN) | NUMBER OF OUTLIERS | NUMBER OF INDIVIDUALS WITH OUTLIER EARNINGS | INDIVIDUALS WITH OUTLIER EARNINGS (%) |
|---|---:|---:|---:|
| $10,000 | 1695 | 334 | 7.38 |
| $15,000 | 269 | 85 | 1.87 |
| $20,000 | 93 | 26 | 0.57 |

- **Trends of average earnings and percent employed by quarter:** [Insert text to note changes in percent employed (with earnings) and average earnings (which include $0s for those who didn't earn in each quarter), and whether the variations from quarter to quarter are expected/reasonable.]

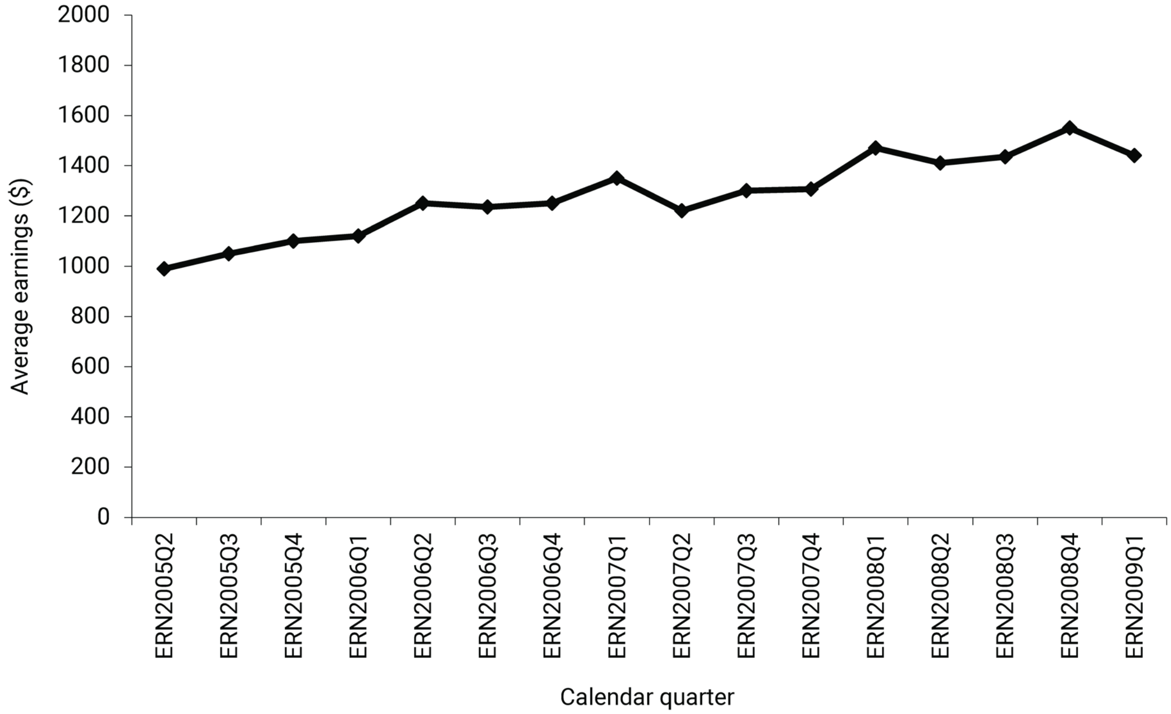
  *Note: The two figures below, Figure 3.2.1 and Figure 3.2.2, are examples of how you can easily see trends in employment and earnings over time. The reasonableness of the trend you see will depend on your study sample and what you are analyzing. For example, when we use UI data to evaluate a training program, we often see an increase in employment and earnings in the quarters following participation.*

  *Note: The UI wage data file is a person-level file. Linking these data to a case-level file will require further data processing and quality control checks.*

**Figure 3.2.1 Percent with Earnings, 2005Q2 through 2009Q1**



## Figure 3.2.2 Average Earnings (in dollars), 2005Q2 through 2009Q1



# Phase 4

# Analyzing the Data

Once the data is prepared and cleaned, your team chooses an analytics method suitable for both your data and your research question. When conducting a data analytics project, selecting the most appropriate method is crucial to effectively answering your research question. For every research question, there are choices about which method will be most effective at answering it. Some research questions are about understanding or describing an issue or dynamic, informing a problem statement or assessing the magnitude of a problem. For example, how many TANF program participants complete training?

In other cases, you might have decided on a dynamic or problem and you want to identify causes and consequences of that dynamic or problem. For example, some TANF participants seem to cycle on and off the TANF program, and on and off employment. What is causing this dynamic known as churning?

Or, you may already have an idea or hypothesis that you want to test for how effectively it addresses the problem or affects the dynamic. For example, does the new coaching approach to delivering services improve the chances that TANF participants get good jobs?

The discussion to define your research question and the analytic approach can contribute to a shared understanding of the primary goals of the data analytics project. **Tool 4.1 – Analyzing Data: Research Questions and Methods** offers examples of questions and approaches that can answer them.

**Tool 4.2 – Did it Work? Interpreting Study Findings** is designed to assist in interpreting findings from three common nonexperimental research designs for estimating the effects of an intervention: pre-/post-test, interrupted time series (ITS), and comparative interrupted time series (CITS). These designs are “nonexperimental” because they do not involve random assignment. These designs can be used when you can observe how outcomes change over time.

Some research questions address what works best: Which intervention or strategy achieved the desired outcome? These efforts can lead to making causal claims. For example, if the analysis shows that program participants who attended job club were employed at a faster rate than those program participants who did not attend job club, you might be tempted to claim that the job club improved employment outcomes. But before making such a claim, it is important to interrogate your analytics results to make sure that they can support your causal claim. For example, it is not enough to simply compare those who received the job club to those who did not receive the job club and attribute the difference to the job club itself. Part of interrogating your results is assessing whether there are plausible alternative explanations for your findings. Any plausible alternative explanations may threaten your analysis's internal validity, which is the extent to which you can be confident that the cause-and-effect relationship claimed in your study does not have alternative explanations. If internal validity is weak or threatened, it means that you can't make a causal claim about whether your intervention achieved the desired outcome.

To support staff members who are learning concepts, vocabulary, and ways to design and conduct program evaluations and different types of evaluations, see **Tool 4.3 – Program Evaluation Resources** and **Tool 4.4 – Evaluation Glossary**. They provide a glossary of terms and links to additional resources for staff.

# Tool 4.1 – Analyzing Data: Research Questions and Methods

For every research question, there are choices about which method will be most effective at answering it.

The discussion to define your research question and the analytic approach can contribute to a shared understanding of the primary goals of the data analytics project. Table 4.1.1 offers examples of questions and approaches that can answer them.

**Table 4.1.1 Sample Questions and Approaches That Can Answer Them**

| IF YOU WANT TO: | YOU NEED A QUESTION AND ANALYTIC APPROACH TO ADDRESS: | TANF PROGRAM EXAMPLE |
|---|---|---|
| Understand typical experience/outcome | Central tendency | What percentage of TANF program participants who leave the program return within [X] months?<br>For TANF customers who return after leaving, how many months, on average, are customers off TANF before returning to cash assistance? |
| Describe or quantify range in experience/ outcome or identify outliers | Dispersion/variation | How do churn probabilities vary across counties/regions? Across customers w/different characteristics?[a] |
| Describe change over time or predict change in the future | Trends, changes | How do churn probabilities vary across time/year? |
| Understand relationships between two or more measures, or cause and effect | Association, causes | Is participation in particular services associated with lower churn rates? |
| Consider differences within and between units (for example, customers within case managers, case managers within counties/regions) | Multilevel analysis | How do churn rates vary among customers served by the same case managers?<br>How do churn rates vary across case managers? |

Note: [a] "Churn" refers to when benefits recipients cycle unnecessarily or unproductively out of and back into a public benefits program. See Rosenbaum, Dottie. 2015. "Lessons Churned: Measuring the Impact of Churn in Health and Human Services Programs on Participants and State and Local Agencies." Website: https://www.cbpp.org/research/lessons-churned-measuring-the-impact-of-churn-in-health-and-human-services-programs-on.

# Tool 4.2 – Did it Work? Interpreting Study Findings

Data analytics projects often seek to answer the question of what works best, or did an intervention or strategy improve a desired outcome. These efforts can lead project teams to make causal claims. For example, if the analysis shows that program participants who attended job club were employed at a faster rate than those program participants who did not attend job club, you might be tempted to claim that the job club improved employment outcomes. But before making such a claim, it is important to interrogate your analytics results to make sure that they can support your causal claim. For example, it is not enough to simply compare those who received the job club to those who did not receive the job club and attribute the difference to the job club itself. Part of interrogating your results is assessing whether there are plausible alternative explanations for your findings. Any plausible alternative explanations may threaten your analysis's internal validity, which is the extent to which you can be confident that the cause-and-effect relationship claimed in your study does not have alternative explanations. If internal validity is weak or threatened, it means that you can't make a causal claim about whether your intervention achieved the desired outcome.

This tool is designed to assist in interpreting findings from three common nonexperimental research designs for estimating program outcomes: pre-/post-test, interrupted time series (ITS), and comparative interrupted time series (CITS) designs. These designs are nonexperimental designs since they do not involve random assignment and are typically used when there is data on outcomes over time. This tool does not suggest when these research designs are appropriate to use or how to use but offers basic information on each.

In most cases, these research designs are not set up in a way that will allow you to make a causal claim. Simply looking at outcomes before and after an intervention is rarely enough to demonstrate a cause/effect relationship. However, the designs described below have improved in recent years and smart implementation of design features can greatly improve the rigor. Lastly, these designs are presented in order of least to most complexity and rigor with pre-/post-test being the least to CITS being the most.

*Instructions: Outlined below is a brief definition of the study design, a sample graph and a checklist to use to assess internal validity for each type of research design followed by a sample graph of findings.*

## Pre-/Post-Test Design

Pre-/post-test designs involve comparing outcomes for individuals at two different points in time—before and after an intervention or policy change is implemented—to see if there is a change in the outcome.[1] For example, programs may compare the average earnings of their

1. Thiese (2014).

participants prior to entering the program to their average earnings after they complete the program. This can be helpful to describe the outcomes of individuals in a program (for example, the average earnings of individuals who attend the program), but it does not provide information on the effectiveness of a program at improving an outcome.[2]

### Sample Graph

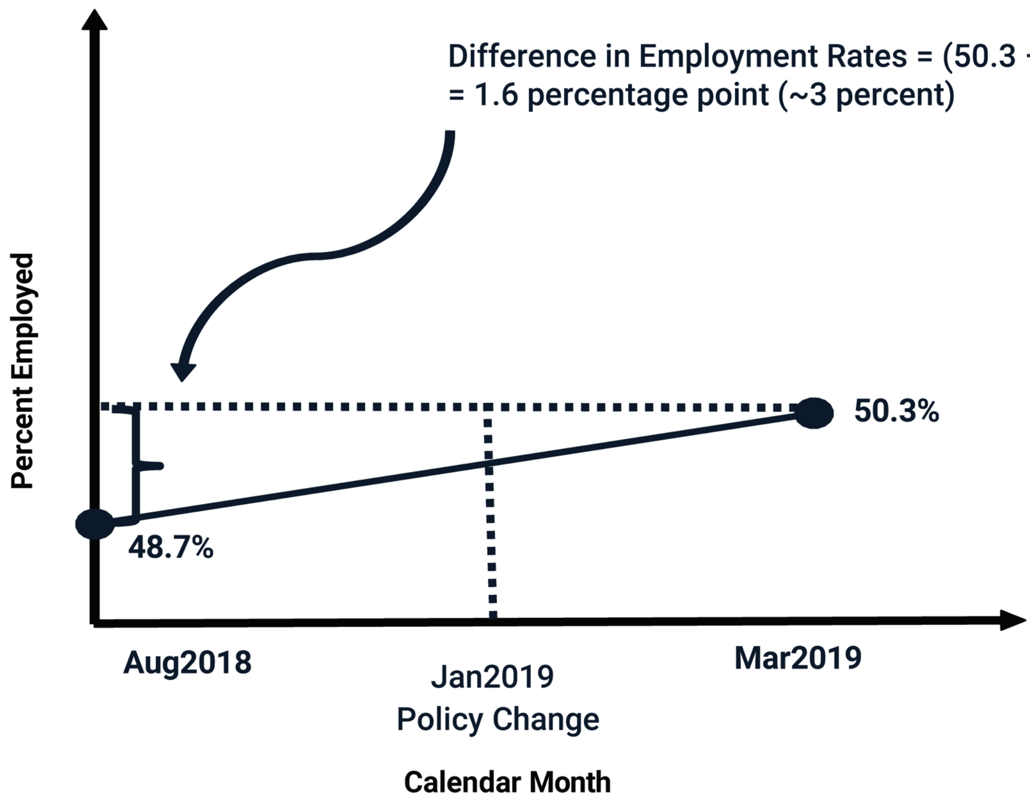
In order to increase employment among TANF recipients, the State of Sufficia made two changes to its CCDF-childcare funded services in 2019:

1. Prioritized TANF clients for childcare services

2. Waived co-payments for TANF families

To evaluate these program changes, the State of Sufficia sought to compare the average employment rate of TANF recipients before and after the two changes were made in 2019. Sufficia found a 1.6 percentage point increase in the employment rate between August 2018 and March 2019, as illustrated in Figure 4.2.1. It should be noted that this type of design does not allow Sufficia to say that the policy changes led to this increase in employment, as it does not account for other potential reasons the employment rate might have increased.

**Figure 4.2.1 Employment Rate Before and After Sufficia Policy Changes**




2. Gopalan, Rosinger and Bin Ahn (2014).

### Validity Checklist

There are several questions to ask when considering or implementing a pre-/post-test design. The answers to these questions can help inform whether you should consider doing a pre-/post-design, and help you better understand the results you see and the claims you can make about them. As shown in Table 4.2.1, focus on whether these potential threats are *plausible*, and not just *possible*.

**Table 4.2.1 Pre-/Post-Test Design: Validity Checklist**

| VALIDITY THREAT | PLAUSIBLE? | ADDRESSED? |
|---|---|---|
| History: Are other events occurring concurrently (for example, a new employer opened in the local area)? | | |
| Maturation: Is there naturally occurring change in the outcome over time (for example, people's earnings tend to increase as they accumulate work experience)? | | |
| Selection: Do characteristics of who is in the sample used to measure the outcome vary by time point? | | |
| Attrition: Are study subjects dropping out? Are you missing outcome data for some people? | | |
| Testing: Are behavior changes occurring due to the act of measuring the outcome (for example, staff are more aware, so they are entering data more consistently)? | | |
| Instrumentation: Did the outcome measure/way the outcome is collected change over time? | | |

## Interrupted Time Series (ITS)

Interrupted time series research designs involve collecting data on outcomes at multiple time points, both before and after an intervention or policy change is implemented, to assess whether there is a difference in the "trend" of the outcome after the implementation of the intervention or policy change (compared to before). Or in other words, whether the intervention or policy change "interrupts" the existing trend in the outcome.[3]

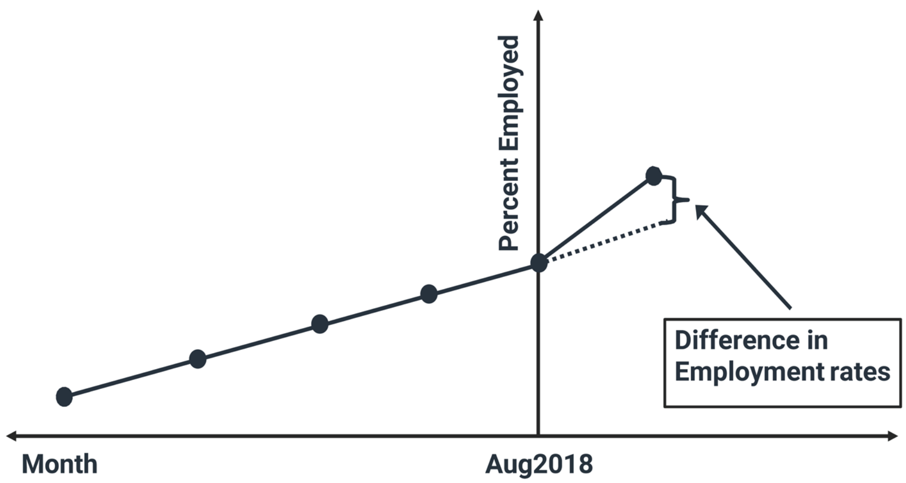
This design is more rigorous than a pre-/post-test design, as having outcome data in multiple time periods can help you see changes that are occurring outside of the intervention or policy change. However, it still does not always allow you to say with confidence that a program led to an effect on an outcome, as it does not control for other outside factors that could also influence the outcome.

3. St. Clair, Hallberg & Cook (2014).

## Sample Graph

The state of Sufficia collected data on employment every month in the year before and in the month after implementing the two program changes mentioned above. The monthly employment rates are shown in Figure 4.2.2. The vertical line shows when the program changes were implemented (in August 2018).

**Figure 4.2.2 Employment Rate at Multiple Points Before and After Policy Change**




Sufficia discovered that the employment rate observed following the program changes deviated from the employment rate trend in the months prior to the changes; the increase was greater than the projected increase in employment. This is seen by the actual employment rate being higher than the line showing the projected employment rate. It should be noted that this type of design does not allow Sufficia to say with full confidence that the program changes led to this increase in employment, unless they have accounted for all other potential reasons the employment rate might have increased.

## Validity Checklist

There are several questions to ask when considering or implementing an ITS design. In general, ask yourself: Does the ITS study address these potential alternative explanations for changes in outcomes? As shown in table 4.2.2, focus on whether these potential threats are *plausible*, and not just *possible*. For ITS studies, look for abrupt *changes* in these aspects *at the time of the intervention*.

### Table 4.2.2 Interrupted Time Series: Validity Checklist

| VALIDITY THREAT | PLAUSIBLE? | ADDRESSED? |
|---|---|---|
| History: Are other events occurring concurrently (for example, a new employer opened in the local area)? | | |
| Instrumentation: Did the outcome measure/way the outcome is collected change over time? | | |
| **Special issue with multiple cross-sectional time series** | | |
| Selection: Do characteristics of who is in the sample used to measure the outcome vary by time point? | | |
| **Special issue with panel/longitudinal time series** | | |
| Maturation: Is there naturally occurring change in the outcome over time (for example, people’s earnings tend to increase as they accumulate work experience)? | | |
| Attrition: Are study subjects dropping out? Are you missing outcome data for some people? | | |
| Testing: Are behavior changes occurring due to the act of measuring the outcome (for example, staff are more aware, so they are entering data more consistently)? | | |

## Comparative Interrupted Time Series (CITS)

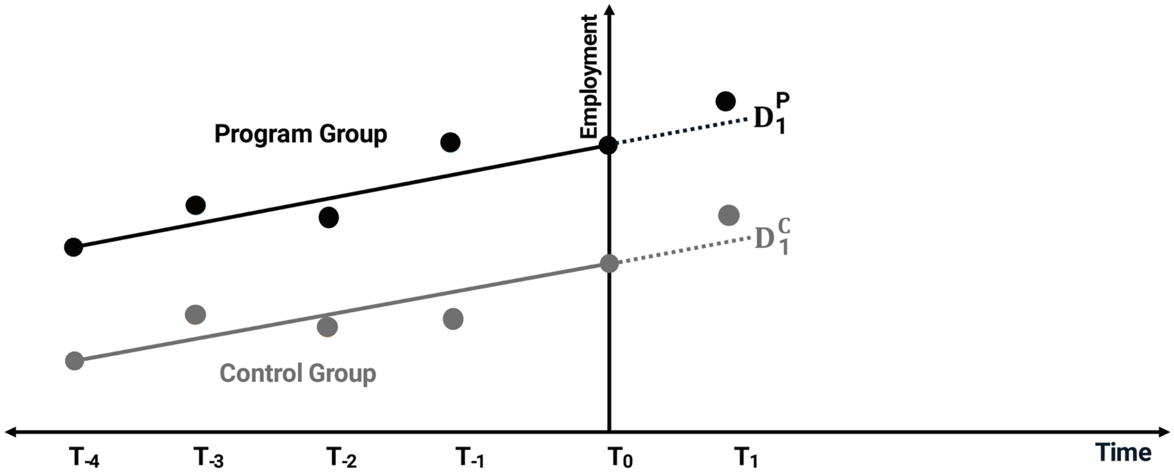
Similar to the ITS design, the CITS design involves collecting outcome data at multiple time points before an intervention and during at least one time point after an intervention. The difference is in this design, outcome data is also collected for a group of individuals that did not experience the intervention or policy at the same time. This inclusion of a “control” or “comparison” group helps you see what would have happened in the absence of the intervention, and whether any outside changes affected the outcomes for those individuals. The CITS design allows for a more robust evaluation.[4]

### Sample Graph

Sufficia decided to compare the employment rates of two groups of participants: one that was subject to the policy changes described above and one that was not. Figure 4.2.3 illustrates the employment outcomes for the program group (the group that was subject to the policy changes, shown in black) and for the control group (the group that did not receive the intervention, shown in gray) at multiple time points before the policy changes and at one point after the policy changes.

4. St. Clair, Hallberg, and Cook (2014).

## Figure 4.2.3 Estimated Effect of the TANF Policy Changes on the Employment Rate




Note: DIP and DIC are the deviation from the baseline trend for the program group and control group, respectively.

Sufficia was able to estimate the effect of the intervention by taking the difference in the deviation from the employment rate trends between the two groups after the intervention (in time = T1).

### Validity Checklist

There are several questions to ask when considering or implementing the CITS design. In general, ask yourself: does the CITS study address these potential alternative explanations for any effects that it finds? As shown in Table 4.2.3, focus on whether these potential threats are *plausible*, and not just *possible*.

## Table 4.2.3 Comparative Interrupted Time Series: Validity Checklist

| VALIDITY THREAT | PLAUSIBLE? | ADDRESSED? |
|---|---|---|
| Selection: Are the average characteristics of individuals in the program and control groups different? | | |
| **Do any of the following affect one research group *differently than the other*?** | | |
| History: Do the events occurring concurrently with the intervention differ (for example, another policy change made by a different agency)? | | |
| Maturation: Does the naturally occurring change in the outcome over time differ (for example, people's earnings tending to increase as they work more)? | | |
| Selection: Do characteristics of who is in the sample used to measure the outcome vary by time point? | | |
| Regression artifacts: That is, the tendency of extreme scores to gravitate toward the mean. | | |
| Attrition: Does the rate of study subjects dropping out differ? Does the rate of missing outcome data differ? | | |
| Instrumentation: Do any changes in the way the outcome is measured/collected differ? | | |

# Tool 4.3 – Program Evaluation Resources

This tool includes selected references to relatively accessible information about various program evaluation topics. The list is neither exhaustive nor intended to include all foundational references to particular topics. Instead, it aims to provide additional "how-to" information and beginning points for further exploration.

A resource that aims to strengthen program managers' understanding of and readiness for program evaluation is "The Program Manager's Guide to Evaluation," published by the Office of Planning, Research, and Evaluation (OPRE) within the Administration for Children and Families. It explains what program evaluation is, its importance, and different steps in the evaluation process, including how to engage an evaluation team, prepare for and design an evaluation, gather credible evidence and analyze data, and share lessons learned.

In addition, OPRE periodically organizes meetings to convene scientists and research experts to advance critical topics in social science research methodology. The meetings provide an opportunity to discuss how innovative methodologies can be applied to policy-relevant questions and help to ensure that government-supported research represents the most scientifically rigorous approaches available. Additional resources on some of the topics listed below can be found under Past Meetings on the OPRE site.

### Topics

- Difference in Differences
- Effect Sizes
- Evaluation Design—General
- Interrupted Time Series (ITS) and Comparative Interrupted Time Series (CITS)
- Matching
  - General
  - Propensity Score Methods
  - Synthetic Comparison Methods
- Multiple Hypothesis Testing
- Null Results
- Regression Discontinuity (RD)
- Subgroups
- Theory of Change and Logic Models

### Difference in Differences

Somers, Marie-Andrée, Pei Zhu, Robin Jacob, and Howard Bloom. 2013. "The Validity and Precision of the Comparative Interrupted Time Series Design and the Difference-in-Difference Design in Educational Evaluation," MDRC Working Paper on Research Methodology. New York: MDRC.

Wing Cody, Kosali Simon, and Ricardo A. Bello-Gomez. 2018. "Designing Difference in Difference Studies: Best Practices for Public Health Policy Research." *Annual Review of Public Health* 39: 453-469.

Bloom, Howard S., Charles Michalopoulos, Carolyn J. Hill, and Ying Lei. 2002. "Can Nonexperimental Comparison Group Methods Match the Findings from a Random Assignment Evaluation of Mandatory Welfare-to-Work Programs?" MDRC Working Paper on Research Methodology. New York: MDRC.

### Effect Sizes

Bloom, Howard S. 1995. "Minimum Detectable Effects: A Simple Way to Report the Statistical Power of Experimental Designs." *Evaluation Review* 19, 5: 547-556.

Bloom, Howard S., Carolyn J. Hill, Alison Rebeck Black, and Mark W. Lipsey. 2008. "Performance Trajectories and Performance Gaps as Achievement Effect-Size Benchmarks for Educational Interventions." New York: MDRC.

Cohen, Jacob. 1988. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates.

Durlak, Joseph A. 2009. "How to Select, Calculate, and Interpret Effect Sizes." *Journal of Pediatric Psychology* 34, 9: 917-928.

Hill, Carolyn J., Howard S. Bloom, Alison Rebeck Black, and Mark W. Lipsey. 2008. "Empirical Benchmarks for Interpreting Effect Sizes in Research." *Child Development Perspectives* 2, 3: 172-177.

Lipsey, Mark W., Kelly Puzio, Cathy Yun, Michael A. Herbert, Kasia Steinka-Fry, Mikel W. Cole, Megan Roberts, Karen S. Anthony, and Matthew D. Busick. 2012. *Translating the Statistical Representation of the Effects of Education Interventions into More Readily Interpretable Forms*. *(NCSER 2013-3000)*. Washington, DC: National Center for Special Education Research, Institute of Education Sciences, U.S. Department of Education.

Angrist, Joshua D. and Jörn-Steffen Pischke. 2015. *Mastering Metrics: The Path from Cause to Effect*. Princeton, NJ: Princeton University Press.

### Evaluation Design—General

El Mallah, S., Gutuskey, L., Hyra, A., Hare, A., Holzwart, R., & Steigelman, C. 2022. "The Program Manager's Guide to Evaluation" (OPRE Report 2022-208). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation.

Murnane, Richard J., and John J. Willett. 2010. *Methods Matter Improving Causal, Inference in Educational and Social Science Research*. Oxford, England: Oxford University.

Shadish, William R., Thomas D. Cook, and Donald T. Campbell. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton, Mifflin and Company.

Bloom, Dan. 2018. "How Is Random Assignment Like a Frying Pan?" *MDRC Reflections on Methodology* (blog), November. Website: https://www.mdrc.org/work/publications/how-random-assignment-frying-pan.

Duflo, Esther, Rachel Glennerster, and Michael Kremer. 2008. "Using Randomization in Development Economics Research: A Toolkit." Pages 3895-3962 in *Handbook of Development Economics*. Amsterdam: Elsevier.

### Interrupted Time Series (ITS) and Comparative Interrupted Time Series (CITS)

Bloom, Howard S. 2003. "Using 'Short' Interrupted Time-Series Analysis To Measure The Impacts Of Whole- School Reforms: With Applications to a Study of Accelerated Schools." *Evaluation Review* 27, 1: 3-49.

Somers, Marie-Andrée, Pei Zhu, Robin Tepper Jacob, and Howard Bloom. 2013. "The Validity and Precision of the Comparative Interrupted Time Series Design and the Difference-in-Difference Design in Educational Evaluation." New York: MDRC.

Wing, Coady, Kosali Simon, and Ricardo A. Bello-Gomez. 2018. "Designing Difference in Difference Studies: Best Practices for Public Health Policy Research." *Annual Review of Public Health* 39: 453-469.

Shadish, William R., Thomas D. Cook, Donald T. Campbell, and C. S. Reichardt. 2002. "Experimental and Quasi-Experimental Designs for Generalized Causal Inference," Book review. *Social Service Review* 70, 3: 510–514.

Tuttle, Christina Clark, Brian Gill, Philip Gleason, Virgina Knechtel, Ira Nichols-Barrer, and Alexandra Resch. 2013. "KIPP Middle Schools: Impacts on Achievement and Other Outcomes, Final Report." Washington, DC: Mathematica Policy Research, Inc.

Wong, Manyee, Thomas D. Cook, and Peter M. Steiner. 2009. "No Child Left Behind: An Interim Evaluation of Its Effects on Learning Using Two Interrupted Time Series Each With Its Own Non-Equivalent Comparison Series." Working Paper 09-11, 18 (7). Evanston, IL: Institute for Policy Research, Northwestern.

## Matching

### *General*

Stuart, Elizabeth A. 2010. "Matching Methods for Causal Inference: A Review and Look Forward." *Statistical Science* 25, 1: 1-21.

### *Propensity Score Methods*

Guo, Shenyang. and Mark W. Fraser. 2014. *Propensity Score Analysis: Statistical Methods and Applications*. Sage Publications: Thousand Oaks, CA.

Stuart, Elizabeth A. 2011. "The Why, When, and How of Propensity Score Methods for Estimating Causal Effects." Unpublished Paper. Fairfax, VA: Society for Prevention Research.

Randolph, Justus J., Kristina Falbe, Austin Kureethara Manuel, and Joseph L. Balloun. 2019. "A Step-by-Step Guide to Propensity Score Matching in R." *Practical Assessment, Research, and Evaluation* 19, 18.

Heinrich, Carolyn, Alessandro Maffioli, and Gonzalo Vazquez. 2010. *A Primer for Applying Propensity-Score Matching*. Washington, DC: Office of Strategic Planning and Development Effectiveness, Inter-American Development Bank.

Caliendo, Marco, and Sabine Kopeinig. 2008. "Some Practical Guidance for the Implementation of Propensity Score Matching." *Journal of Economic Surveys* 22, 1: 31-72.

### *Synthetic Comparison Methods*

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *Journal of the American Statistical Association* 105: 490, 493-505.

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2014. "Comparative Politics and the Synthetic Control Method." *American Journal of Political Science* 59, 5: 495-510.

Cook, Thomas D., William R. Shadish, and Vivian Wong. 2008. "Three Conditions Under Which Experiments and Observational Studies Produce Comparable Causal Estimates: New Findings from Within-Study Comparisons." *Journal of Policy Analysis and Management* 27, 4: 724-750.

## Multiple Hypothesis Testing

Schochet, Peter M. 2008. *Technical Methods Report: Guidelines for Multiple Testing in Impact Evaluations*. Washington, DC: Institute of Education Sciences, U.S. Department of Education.

### Null Results

Herrington, Carolyn D., Rebecca Maynard. 2019. "Editors' Introduction: Randomized Controlled Trials Meet the Real World: The Nature and Consequences of Null Findings." *Educational Researcher* 48, 9: 577-579.

Jacob, Robin T., Fred Doolittle, James Kemple, and Maire-Andree Somers. 2019. "A Framework for Learning From Null Results." *Educational Researcher* 48, 9: 580-589.

Landis, Ronald S., Lawrence R. James, Charles E. Lance, Charles A. Pierce, and Steven G. Rogelberg. 2014. "When is Nothing Something?" Editorial, Null Results Special Issue, *Journal of Business and Psychology* 29, 2: 163-167.

### Regression Discontinuity (RD)

Bloom, Howard S. 2012. "Modern Regression Discontinuity Analysis." *Journal of Research on Educational Effectiveness* 5, 1: 43-82.

Imbens, Guido W. and Thomas Lemieux. 2008. "Regression Discontinuity Designs: A Guide to Practice." *Journal of Econometrics* 142, 2:615–635, 2008.

Jacob, Robin, Pei Zhu, Marie-Andree Somers, Howard Bloom. 2012. "A Practical Guide to Regression Discontinuity." New York: MDRC.

Howell, Sabrina T. 2016. "Financing Innovation: Evidence from R&D Grants." Unpublished Paper. New York: New York University.

### Subgroups

Bloom, Howard S., Charles Michalopoulos. 2013. "When is the Story in the Subgroups?" *Prevention Science* 14, 179–188.

### Theory Of Change and Logic Models

Bangser, Michael. 2014. "A Funder's Guide to Using Evidence of Effectiveness in Scale-Up Decisions." New York: MDRC.

Epstein, Diana and Jacob Alex Klerman. 2012. "When is a Program Ready for Rigorous Impact Evaluation? The Role of a Falsifiable Logic Model." *Evaluation Review* 36, 5: 375-401.

W.K. Kellogg Foundation. 2004. "Logic Model Development Guide: Using Logic Models to Bring Together Planning, Evaluation, and Action." Battle Creek, MI: W.K. Kellogg Foundation.

# Tool 4.4 – Evaluation Glossary

This tool provides a glossary of terms for staff members who are learning concepts and vocabulary related to program evaluation.

### Propensity Score Matching

- **Propensity score:** the probability of being in the group that is offered the program or affected by the policy change based on their observed characteristics.

- **Propensity score matching (PSM):** a method for identifying a comparison group that has observed characteristics similar to those of the program group.

### Counterfactual

- What happens in the absence of a program or policy change? In an experiment, the control group provides the **counterfactual**. Without an experiment, we can use other methods (like propensity score matching) to find a counterfactual (often called a comparison group).

### Study Types and Definitions

- **Outcome:** The level on a measure for an individual or a group of individuals that is used to assess program performance.

- **Impact:** The difference between the average outcomes for the program group versus the comparison or control group. Outcomes and impacts are not the same.

- **Pre-/post-test:** A research design in which the same outcome is assessed before and after a program. This is a weak research design because typically many other factors could explain pre/post differences such as maturation, other historical events, and so on.

- **Longitudinal tracking study:** Follows study participants over time and collects data to measure their outcomes. While this can be useful for generating hypotheses, it is weak for the purposes of causal inference.

- **Regression discontinuity design:** Researchers take advantage of a threshold in the program eligibility criteria (for example, a test score or income threshold). Individuals above (or below) the threshold serve as the program group and individuals below (or above) the threshold serve as the comparison group. The estimated impact is defined only for individuals very close to the threshold. The validity of the design assumes that at the threshold, the design is equivalent to a random assignment design. This is considered the second most rigorous research design. But it requires more sample than a randomized controlled trial (RCT).

- **Comparative interrupted time series:** Uses longitudinal data for a program group and a matched comparison group to estimate the effects of an intervention. The analysis compares the two groups' deviations from their baseline trends after the intervention. Because of the comparison group and the multiple observations, this method is more rigorous than pre-/post- tests.

- **Synthetic controls:** A statistical method used to evaluate the effect of an intervention that affected a place. The comparison group is similar to a propensity score created group but other design elements such as matching on time trends are included.

- **Random assignment:** Divides study participants into a "treatment group" (or "program group") that is eligible to receive program services and a "control group" that is not eligible. Comparing the outcomes of the groups over time allows us to estimate the impacts of the program. This is the most rigorous research design.

### Selection Bias

- The bias introduced by the selection of individuals, groups, or data for analysis in such a way that proper randomization is not achieved, thereby failing to ensure that the sample obtained is representative of the population intended to be analyzed.

### Observed and Unobserved Characteristics

- PSM matches individuals in the program and comparison groups on **observed characteristics** – so the two groups are "balanced." This may account for some of the differences in **unobserved characteristics**. And this balancing of characteristics can help reduce **selection bias**.

- **Observed characteristics** are variables for which you have measurements in your dataset.

    - Baseline characteristics, for example, age, gender, employment history

- **Unobserved characteristics** are variables for which you *don't* have measurements in your dataset.

    - Motivation, grit, decision making, for example

### Comparison Pool and Comparison Group

- A **comparison pool** is identified first. Then, you match your program group to the larger pool to identify your comparison group. The **comparison group** is a subset of the comparison pool.

- When creating your comparison pool:

- The comparison pool should be larger than program group.
- You should refine your comparison pool as much as you can at this stage.
- You need good data on the comparison pool to help with selection bias and unobserved characteristics.

- Questions to consider when identifying and refining your comparison pool:
  - Can a valid comparison group be identified?
  - Are the key characteristics of individuals measurable?
- The comparison group should ideally share the following characteristics with the program group:
  - Same geographic location
  - Same time period
  - Meets the program eligibility criteria
  - Same data available

### Calculating Propensity Scores

- The propensity score is the probability of being in the group that is offered the program or policy change based on their observed characteristics. This is often calculated using logistic regression. **Logistic Regression** is used when the dependent variable (target) is categorical.
- What characteristics should you use to calculate propensity scores?
  - You want to balance the program and comparison groups on many characteristics.
  - Characteristics must be measured before the program or policy change ("baseline").
  - Characteristics to consider including:
    - Baseline measures of the outcome
    - Demographics
    - Other characteristics that predict the outcome
    - Interactions and higher-order terms

### Common Support

- We want a distribution of propensity scores – not everyone should have the same probability of being offered the program or policy change. In lay terms, common support means that you have enough similar individuals in your comparison pool to match. It becomes a problem at the extremes of a distribution (for example, if matching the individuals with the very lowest income, or the most serious barriers to employment)

- **Minima/maxima**: Drop observations where scores are outside the range of the other group.

- **Trimming**: Require a specific percentage of observations within minima and maxima.

### Matching

- Individuals in the program group are matched to individuals in the comparison group on their calculated propensity scores. After matching, it is critical to check that the two groups are similar on observed characteristics.

- **Considerations**

  - Replacements: Can individuals in the comparison pool be matched to more than one person?

  - Oversampling: Do you want to match individuals to multiple comparison pool members?

- **Methods**
  *The choice of matching method is not as important as having good data to match on!*

  - **Nearest neighbor:** The most straightforward matching estimator is nearest neighbor (NN) matching. The individual from the comparison group is chosen as a matching partner for a treated individual that is closest in terms of propensity score.[1]

  - **Caliper:** Closest match within a specified boundary (consider this like a range beyond which you would not consider an individual a match).[2]

---

1. Caliendo, Marco Sabine Kopeinig (2005).
2. Caliper and radius matching: NN matching faces the risk of bad matches if the closest neighbor is far away. This can be avoided by imposing a tolerance level on the maximum propensity score distance (caliper). Imposing a caliper works in the same direction as allowing for replacement. Bad matches are avoided and hence the matching quality rises. However, if fewer matches can be performed, the variance of the estimates increases. Applying caliper matching means that the individual from the comparison group is chosen as a matching partner for a treated individual that lies within the caliper ("propensity range") and is closest in terms of propensity score.

- **Stratification:** grouping by score range.
- **Kernel matching (KM) and local linear matching (LLM):** non-parametric matching estimators that use weighted averages of all individuals in the control group to construct the counterfactual outcome.

- **Assessing the match**
  *You want to confirm that the program and comparison group are similar on characteristics measured before the program or policy change.*
  - Are the average characteristics of individuals similar across the two groups? (The preferred answer is “yes.”)
  - Can you predict who is in the program group based on their characteristics? (The preferred answer is “no.”)

# Phase 5

# Communicating the Results

This phase includes the final steps for formatting and sharing findings from the analysis, offering the opportunity to translate what was learned and why it matters to different audiences. There are many ways to communicate the findings along with their implications and the methods and data used to complete the analyses. Similarly, there are a variety of audiences who may benefit from understanding and learning from your findings. This means that determining your audience informs the way you will communicate, and customization of the materials is important to effectively achieve your objective.

**Tool 5.1 – Project Summary Report Template** outlines a table of contents in a typical research report that can serve as the final output of a data analytics project. It can draw from a variety of other outputs created during the project itself such as the project scope, analysis plan, or interim reports. The tool offers six categories of report content with prompting questions for you to consider along with possible appendices such as code used during the analyses.

A final report is one type of dissemination product that can be used to share the lessons and insights learned from conducting a data analytics project. Another type of product is a verbal presentation or briefing for project sponsors, managers, executives, and external partners.

Briefing internal and external audiences is an important step toward securing support for data analytics overall and for identifying champions who value the work and make it possible to continue current projects and queue up future projects. Therefore, planning and conducting briefings is an important milestone that should be included in any data analytics project. The purposes of briefings are to (1) strengthen support among those interested in the data analytics work that you are doing in your project, and (2) generate ideas and demand for the next set of questions or next project.

**Tool 5.2 – Briefing Instructions and Template** offers a five-part table of contents to guide your preparation for a verbal presentation or briefing. The template invites you to prepare how you are telling the story of the project, who is your target audience, what is the purpose of the briefing, and what is your message.

# Tool 5.1 – Project Summary Report Template

***Instructions:*** A final report is one type of dissemination product that can be used to share the lessons, insights, and steps to conducting a data analytics project. It can inform decisions and discussions and provide a record for future reference. Below is a suggested table of contents listing key topics that can be included in a final project summary report. Please note, several sections of the report can leverage content from other project documents (for example, **Tool 1.2, Project Scope: Instructions and Template**, in Phase 1).

**Background Documents:** You should draw from any of these existing documents to write your Project Summary Report:

1. Literature review

2. Project scope

3. Analysis plan

4. Prior summary report drafts or memos

5. Presentations

**Suggested Table of Contents:**

1. Research questions
    a. Succinct statement
    b. Why does it matter/how do you know it will be used by, for example, administrators or policymakers?
    c. What policy, programmatic or knowledge gaps are you filling?

2. Data sources
    a. What are they? What information does each data source provide?
    b. How are they combined?
    c. What are the important and notable quality issues?

3. Methods
    a. Which methods did you apply?
    b. How have they been applied?

4. Findings
    a. Descriptive statistics
    b. Graphics
    c. Which findings, if any, inspired, surprised (busted a myth), or discouraged you?

5. Caveats
    a. Scalability to other states (where might it go)
    b. Coverage
    c. Quality (timeliness/local validity)
    d. Inferential validity
    e. Examples
    f. Challenges

6. Next steps and measures of success
    a. Measures of success for your project, such as reaction from your key supporters or champions
    b. Next set of research questions and timeline OR next new project and timeline

Appendices. *These items can be included, if available:*

1. Code
    a. Useful code project team has used
    b. Code project team has developed that's available for reuse
    c. GitHub pointers

2. Visuals

3. Anecdotes/stories

4. Relevant project memos
    a. Quality control
    b. Analysis plan

# Tool 5.2 – Briefing Instructions and Template

Briefing internal and external audiences is an important step to secure support and identify champions who value the work and make it possible to continue current projects and queue up future projects. Therefore, planning and conducting briefings is an important milestone that should be included in any data analytics project. The purposes of briefings are to (1) strengthen support among those interested in the data analytics work that you are doing in your project, and (2) generate ideas and demand for the next set of questions or project. The purpose of this tool is to offer questions for consideration and discussion as you prepare to communicate about your data analytics project.

***Instructions:*** Consider the following questions to complete the template.

**What:** How are you telling the story of your project as you enhance your data analytics capability? How are you demonstrating the value of data and data analytics to various audiences so:

- The potential is appreciated so that demand for analytics continues.
- The resources needed are preserved amid competing demands.
- The findings or insights are used for decision making.

**Who:** Who is your target audience(s)? You identify key individuals whose understanding, and support of your data analytics project will increase the likelihood that it is understood and can continue. Who appreciates the potential of data and data analytics; preserves resources for analytics efforts; uses the insights for decision making? Any of these individuals can be in your target audience.

- Peers
- Supervisors
- Managers, mid-level and senior
- Executive leaders
- External (for example, media, legislature, auditor)

**Purpose:** What is the purpose or what is your intended outcome?

- To discuss and debate
- To request more resources (for example, staff time)
- To inform next steps of a project or the next project (for example, next question or hypothesis)
- To make decisions, types include:
  - fiscal or budget
  - policy or legislative
  - administrative or program rules
  - research and evaluation

**What:** What is your message? This will be customized to the audience whose ongoing support for data analytics and resources to make decisions using data is needed.

- Executive summary of findings and recommendations (for example, major headlines)
- Program and policy implications
- Technical explanation of data sources, code, definitions

**When:** When is the right time to deliver the message to achieve your purpose? The timing of a briefing is most informed by the briefing's purpose and can be dependent on the audience's availability or appetite for the information.

# References

Caliendo, Marco, and Sabine Kopeinig. 2005. "Some Practical Guidance for the Implementation of Propensity Score Matching." Journal of Economic Survey 22,1: 31-72.

Coleridge Initiative. n.d. "Applied Data Analytics Training Programs." New York: Coleridge Initiative. Website: https://coleridgeinitiative.org/applied-data-analytics-training-programs.

Goerge, Robert M., Emily R. Wiegand, and Leah Gjertson. 2021. "Unpacking Data Use in State TANF Agencies." OPRE Report 2021-94. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Goerge, Robert M., Emily R. Wiegand, Emma K. Monahan, and Leah Gjertson. 2022. "Exemplary Data Use by State TANF Agencies." OPRE Report No. 2022-133. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Gopalan, Maithreyi, Kelly Rosinger, and Jee Bin Ahn. 2014. "Use of Quasi-Experimental Research Designs in Education Research: Growth, Promise, and Challenges." Review of Research in Education 44,1: 218-243.

Harris School of Public Policy. 2020. "Harris Policy Labs Project Scope Template." Chicago: University of Chicago.

MDRC. 2023. "TANF Data Collaborative Pilot Profiles: A Collection of Data Analytics Projects from State and County TANF Agencies." Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services Website: https://www.acf.hhs.gov/opre/report/tanf-data-collaborative-pilot-profiles-collection-data-analytics-projects-state-county.

Office of Information and Regulatory Affairs. 2018. "Attachment 1–TANF Data Innovation (TDI) Needs Assessment Survey 2018-08-27." Washington, DC: Executive Office of the President, Office of Management and Budget. Website: https://www.reginfo.gov/public/do/PRAViewIC?ref_nbr=201809-0970-005&icID=233062.

Pincus, Erica. 2018. "Finding the Right Fit: How to Scope a Data Project". Website: https://datasmart.hks.harvard.edu/news/article/finding-right-fit-how-scope-data-project.

Rosenbaum, Dottie. 2015. "Lessons Churned: Measuring the Impact of Churn in Health and Human Services Programs and State and Local Agencies." Website: https://www.cbpp.org/research/lessons-churned-measuring-the-impact-of-churn-in-health-and-human-services-programs-on.

St. Clair, Travis, Kelly Hallberg, and Thomas D Cook. 2014. "Causal Inference and the Comparative Interrupted Time Series Design: Findings from Within-Study Comparisons." Society for Research on Educational Effectiveness. Website: https://eric.ed.gov/?id=ED562724.

Thiese, Matthew S. 2014. "Observational and Interventional Study Design Types; An Overview." Biochemia Medica 24,2: 199–210.

Wavelet, Melissa, Stephanie Rubino, Caroline Morris, and Maryah Garner. 2022. "Increasing Data Analytics Capacity in State TANF Agencies: The TANF Data Collaborative Approach." OPRE Report 2022-124. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services

Wiegand, Emily R., Emma K. Monahan, Rose Geoghegan, Melissa Wavelet, and Robert M. Goerge. 2023. "Strengthening Analytics in Government Agencies: A Toolkit for Sustainable Data Use." OPRE Report 2023-48. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Yang, Edith, Sharon Zanti, T.C. Burnett, Richard Hendra, Dennis Culhane, Zarni Htet, Della Jenkins, Camille Preel-Dumas, and Electra Small. 2022. "Expanding TANF Program Insights: A Toolkit for State and Local Agencies on How to Access, Link, and Analyze Unemployment Insurance Wage Data." OPRE Report 2022-226. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

**NEW YORK**
200 Vesey Street
23rd Flr.
New York, NY 10281
Tel: 212 532 3200

**OAKLAND**
475 14th Street
Suite 750
Oakland, CA 94612
Tel: 510 663 6372

**WASHINGTON, DC**
750 17th Street, NW
Suite 501
Washington, DC 20006

**LOS ANGELES**
11965 Venice Boulevard,
Suite 402
Los Angeles, CA 90066