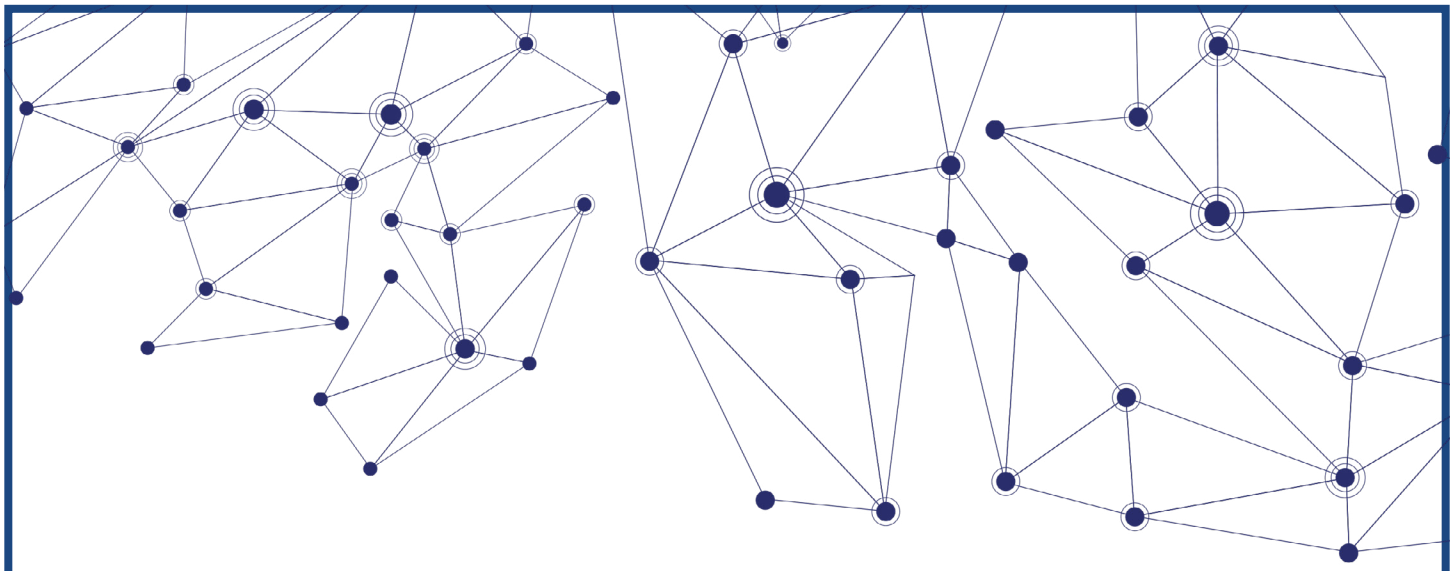


THE BUSY PERSON'S GUIDE TO
MEASURING AND IMPROVING PROGRAM PERFORMANCE

A Step-by-Step Guide to Using Measurement for Better Results

Kevin M. Callahan, PhD

Bloomberg American Health Initiative
Johns Hopkins Bloomberg School of Public Health

CONTENTS

CONTENTS

ACKNOWLEDGEMENTS

This guide would not have been possible without the advice and encouragement of Dr. Elizabeth Stuart, Associate Dean for Education and Professor in the Departments of Mental Health, Biostatistics, and Health Policy and Management. Dr. Stuart's reviews of early drafts of this guide are the primary reason it is more scientifically sound and rigorous than it might have been. Furthermore, without Dr. Stuart's unflagging enthusiasm and support, I doubt I would have been able to finish it.

I owe a huge debt of gratitude to the five members of my advisory group: Gayne Barlow-Kemper, Thomas Higdon, Jackie Range, Gail Nutter Reid, and Tulha Siddiqi. Their experiences managing public health and human service programs in the Baltimore metropolitan area — and the many insights they have gained from those experiences over the years — had an enormous impact on the content, format, and style of this guide. And because of the amazing work of Mindi Levin, Director of the Student Outreach Resource Center (SOURCE), I could recruit these experienced professionals almost effortlessly from SOURCE's already established partnerships with a large and diverse group of community-based organizations.

I am also grateful for the assistance of Dr. Amirah Wahdi, a student in the MSPH program of the Department of Population, Family & Reproductive Health, and Ms. Katherine Klosek, Senior Implementation Advisor at the Center for Government Excellence (GovEx). Amirah took on the very onerous task of summarizing and synthesizing the health and human service data sets that are referenced in this guide. Kat was a primary source for information about a variety of performance measurement approaches, and she was quick to offer her help in extending the reach of this guide (something of which I knew very little).

Dr. Joshua Sharfstein, Director of the Bloomberg American Health Initiative and Vice Dean for Public Health Practice and Training was instrumental in refining the original concept for this guide, which grew out of meetings of the Bloomberg American Health Initiative's Evidence Working Group during the first few months of 2017. Dr. Sharfstein also conceived the idea of assembling a group of current public health and human service program leaders to advise me on the scope and content of the guide, and he gave me valuable feedback on a near-final draft of it.

And finally, I thank Mark Friedman, founder of the Fiscal Policy Studies Institute and author of *Trying Hard is Not Good Enough* (Friedman, 2005), and Phil Lee, president of Clear Impact (https://clearimpact.com/). Although this guide draws from several different approaches for measuring and improving program performance, it relies most heavily on an approach commonly known as "Results-Based Accountability" (RBA). RBA has its roots in numerous places (Schilder, 1997), but Mark has elaborated and extended it the most, and Phil, a longtime colleague who introduced me to RBA more than a decade ago, has been its most tireless advocate.

WHAT THIS GUIDE IS AND ISN'T

Measuring and improving the performance of health and human service programs is an endeavor that encompasses a lot of areas. Identifying performance measures and collecting and analyzing data for them are certainly big parts of it, but the other half of it — using data to improve performance — sometimes fails to get started. There are at least two explanations for this. First, good performance measurement can be difficult work. It often (but not always) includes several tasks that many people would just as soon avoid (using statistics and acknowledging performance problems are just two examples). Second, performance improvement is very different from performance measurement. The skills that enable a program manager to measure performance well are quite different from the skills that enable him or her to successfully improve that performance, and many program managers only get the opportunity to perfect their skills in one or the other. This skills gap, combined with the tendency to avoid some tasks that are part of the performance measurement and improvement process, prevent some performance measurement projects from ever getting off the ground and others from using the results of measurement to improve performance.

This guide arose, in part, out of an interest in creating a how-to manual that would provide — in one place — proven methods, tools, and advice for measuring and improving program performance. Although other handbooks, manuals, and guides have done the same, this one was designed for the busy professional who hardly has a spare minute. In other words, it aims to be not only complete but also brief and easy to use. In addition, it includes certain tools, methods, and other information that have proved to be quite useful, but other manuals and guides haven't covered sufficiently or at all. Of course, as a guide aiming to be brief, it does not contain as much information about some topics in performance measurement and improvement as other, longer guides do. It is hoped, however, that it provides everything that is necessary and some of the many things that are useful.

INTRODUCTION

Many public health and human service professionals struggle to measure the performance of the programs they manage despite a number of recent, large-scale efforts to assist them.[1] Although they recognize that measuring the performance of their programs can lead to better results (Dilley, Bekemeier, & Harris, 2012; Livingood et al., 2013; Zakocs, Hill, Brown, Wheaton, & Freire, 2015) many public health and human service professionals continue to find performance measurement frustrating, burdensome, and risky. All too often, they discover that

- It doesn't produce information that is useful for making decisions about how to improve or change their programs
- It is too complicated, time-consuming, and expensive
- Their senior leaders, funders, and other stakeholders misinterpret or misuse the performance results.

These views are common, and they demonstrate a critical need for simpler and more effective ways to measure program performance, as well as a need for better ways to use and report the results of such efforts (Bradt, 2009; Derose, Asch, Fielding, & Schuster, 2003; Frieden, 2014; Zakocs et al., 2015).

This guide for the busy (and perhaps apprehensive) person who manages or implements a public health or human service program aims to meet these needs by offering a sensible step-by-step process for measuring program performance, with easily accessible and easy-to-learn tips, tools, and methods for accomplishing each step in this process as efficiently and effectively as possible. In laying out this process, this guide also shows program managers and staff how they can

- Use performance data to identify the most effective ways to improve their programs
- Better define the role and contribution of their programs in improving the lives of the people they serve
- Identify and address hidden assumptions that may be hampering good decisions
- Better communicate and build trust with senior leaders and funders

1 Examples of such initiatives include 100 Million Healthier Lives (http://www.100mlives.org/), the Turning Point Performance Management National Excellence Collaborative (http://www.phf.org/resourcestools/pages/turning_point_project_publications.aspx), and the CDC State, Tribal, Local & Territorial Public Health Professionals Gateway (https://www.cdc.gov/stltpublichealth/index.html).

OVERVIEW OF THIS GUIDE'S APPROACH TO MEASURING PERFORMANCE

Although this guide draws from several different approaches for measuring and improving program performance, it relies most heavily on an approach commonly known as "Results-Based Accountability" (RBA). This approach for measuring program performance is simpler and more practical than other approaches, and it places a high value on figuring out if the people that a program serves are any better off after they participate in it. However, it also works well for measuring specific services and processes, as well as other aspects of a program's performance, including its productivity, efficiency, and quality. Perhaps most importantly, RBA uses information about a program's performance to develop strategies for improving the program and getting the results that the program was designed to bring about.[2]

Evaluations of RBA have found it easy to understand and follow, highly motivating, and very effective in (a) focusing attention on the impact of a program on its clients and (b) acting on information about a program's performance to improve it (Opinion Research Services, 2011). For these reasons (and numerous others), many organizations and agencies have cited RBA as a critical factor in the success of their performance measurement and improvement efforts (Herzog, Cooper, & and Holmes, 2017; Lee, 2013; Macintyre, 2016; Zachary, Brutschy, West, Keenan, & Stevens, 2010).

Over the past 20 years, results-based accountability has been defined in slightly different ways, but it typically includes the following steps (some of which are a part of other performance measurement approaches as well):

The Results-Based Accountability (RBA) Process[3]	
1	Define the program, service, or process that will be measured
2	Identify possible measures of performance and choose the most important ones
3	Collect already existing or new data for these measures
4	Analyze and interpret this data
5	Use the findings — and the story behind them — to identify what might work to improve performance
6	Present and discuss the findings, the story behind them, and what might work to improve
7	Choose and implement the best ideas for improving performance

An effective way of summarizing the results of a process like this one is shown on the next page, and this guide is organized in much the same way.

2 RBA can also be used to measure and improve whole populations and communities, but this guide focuses primarily on its use for measuring and improving individual programs, program services, and program processes — especially those that play a role in improving health or well-being.

3 The RBA process shown here is based on one presented in *Trying Hard is not Good Enough* (Friedman, 2005)

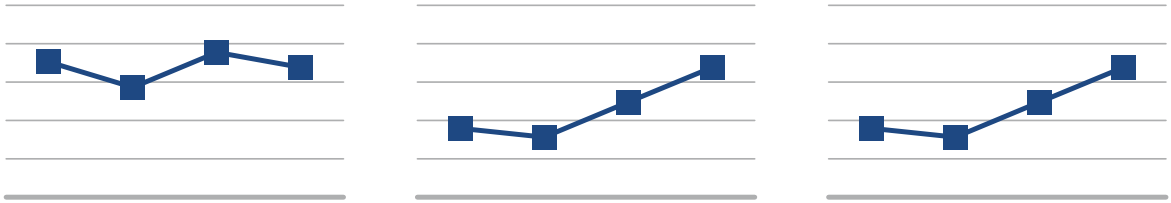
FIGURE 1

Performance Measurement and Improvement Summary Report[4]

Name of Program, Service, or Process Measured:

Purpose and Objectives:

Most Important Measures of Performance:



Story Behind the Data for these Measures:

Partners with a Role to Play:

What Might Work to Improve Performance:

Best Options for Improving Performance:

4 Derived from *Trying Hard is not Good Enough*, Friedman (2005)

STEP 1: DEFINING THE PROGRAM, SERVICE, OR PROCESS THAT WILL BE MEASURED

As with many other performance measurement processes, the first step in the RBA process is to identify the program, service, or process that will be measured. RBA can be used to measure (and improve) the performance of just about anything, including entire organizations and systems. Nevertheless, **those who are new to the RBA process might want to first use it to measure the performance of a single program, service, or process (and a relatively small and simple one at that) before using it to measure the performance of an organization or system, or the peformance multiple services or processes at one time.**

An important part of identifying the program, service, or process that will be measured is clarifying its primary purpose and objectives. The next step in the RBA process — Identifying performance measures — will be easier to accomplish if the purpose and objectives of the program, service, or process that will be measured are fully understood and widely accepted. **If its objectives are fuzzy, or their importance relative to one another is a matter of some debate, one or more of the following strategies may help clarify and get agreement on them:**

- Ask staff which objectives are most important, taking into consideration what clients need and want from the program/service/process
- Create a logic model for the program/service/process (if one does not exist already)[5]
- Examine the objectives for other, similar programs/services/processes

TIP

A good program, service, or process to measure has one or more of these characteristics:

- Well-defined objectives (preferably just a small number of them)
- Well-defined stakeholders/clients/customers
- A key role in enabling an organization to carry out its mission
- An interest in identifying and reporting on the results it is getting
- Staff who support performance measurement

TIP

In the field of performance measurement (and the fields of planning, budgeting, and evaluation), too many terms are used to express the same ideas. As a result, individuals who are just beginning to work together to measure performance can lose a lot of time simply trying to understand and communicate with each other. You can avoid this by reaching agreement at the start on the terms that you and your colleagues will use and then applying these terms consistently. For example, this guide consistently uses *measure* rather than *indicator* , *objective* rather than *goal* , and *result* rather than *outcome* (except when (a) referencing the work of others who use this term and (b) discussing logic models, which often use this term as one of their components).

5 For a good overview of logic models and how to create them, see the Innovation Network's Logic Model Workbook (http://www.pointk.org/resources/files/logic_model_workbook_0.pdf). Also see The funder of my program requires a logic model to measure performance. How can I integrate RBA into a logic model approach?.

KEY POINTS

STEP 1: DEFINING THE PROGRAM, SERVICE, OR PROCESS TO BE MEASURED

The RBA process can be used to measure and improve the performance of just about anything, including entire organizations and systems.

Those who are new to the RBA process might want to first use it to measure the performance of a single program, service, or process before using it to measure the performance of an entire organization or system or multiple services or processes at one time.

An important part of identifying a program, service, or process to measure is clarifying its primary purpose and objectives so they are fully understood and widely accepted. Doing this before moving on to the next step in the performance measurement process — identifying performance measures — will make the latter a lot easier.

Now that you have identified the program, service, or process whose performance you want to measure, your next step is to determine *how* it will be measured. This task is a lot easier if you begin it with three questions. Because these questions encompass the entire universe of performance measures, they provide some order and structure to what could otherwise be a haphazard and even paralyzing process. In addition, they will give you a lot more confidence that you are choosing the right measures of performance for your program.

USING THREE QUESTIONS TO IDENTIFY POSSIBLE PERFORMANCE MEASURES

Three important questions to consider when identifying performance measures are

- **How much did we do?**
- **How well did we do it?**
- **Are our clients any better off?**

Answering these questions is a terrific way to build a comprehensive list of possible performance measures.

	EXAMPLES		
Performance Questions	*Vocational Education*	*Mental Health*	*Juvenile Justice*
HOW MUCH WE DO?	Number of Individuals trained	Number of outpatient client visits	Number of youth in custody
HOW WELL DID WE DO IT?	Percent of trained individuals with improved knowledge/skills	Percent of client intakes or assessments completed on time	Percent of detention facility occupied
ARE OUR CLIENTS ANY BETTER OFF?	Percent of trained individuals who have found and kept a job	Percent of clients with improved symptoms or health	Percent of youth violating probation

As you might have guessed, measuring *"How much did we do?"* is really measuring productivity (and sometimes efficiency or timeliness), while measuring *"How well did we do it?"* is really measuring quality (which can include efficiency, client satisfaction, effectiveness, and more). Although measures that will answer these questions are important and necessary, **the most important measures are those that will answer *"Are our clients any better off?"*** because the purpose of most, if not all, health and human service programs is to improve the health or well-being of the people they serve. Consequently, this question is one that your program's stakeholders, including its funders, are likely to ask (sometimes repeatedly). They will be delighted and impressed if you can give them a conclusive answer!

ADDITIONAL THOUGHTS

THE FUNDER OF MY PROGRAM REQUIRES A LOGIC MODEL TO MEASURE PERFORMANCE. HOW CAN I INTEGRATE RBA INTO A LOGIC MODEL APPROACH?

A considerable number of funding organizations use logic models to plan and manage their programs. A logic model is useful for planning a program because it can assist a program manager in identifying all of the components (that is, the inputs, activities, outputs, and outcomes) of a program and explaining how that program is intended to work. A logic model is also useful for designing and explaining an evaluation of a program.

However, logic models also have several limitations. Logic models that contend that certain program activities will produce certain outputs, or that certain outputs will result in certain outcomes, may be oversimplifying the causes and effects of these program components. This mistake is common in logic models for public health and human service programs that are trying to solve problems that have numerous causes. Furthermore, when a logic model contends that the program elements in it are necessary and sufficient for achieving certain outcomes, it may limit opportunities for identifying alternative ways to achieve these outcomes.

Another limitation of logic models is that they require a certain level of skill to create. This can be an obstacle to collaboration between a program's staff and a program's stakeholders when they want to work together to improve a program's performance or plan the future of a program.

Despite their limitations, logic models can be helpful when introducing RBA to people who are unfamiliar with it. Although RBA and logic models use different terminologies, mapping the results of an RBA process to the more familiar terms of a logic model can help orient people to RBA by showing that it has a lot in common with a logic model[6]. Furthermore, showing a logic model and the results of an RBA process side by side can meet a requirement for a logic model while simultaneously promoting a framework that better supports collaboration and problem solving and better reflects what health and human service programs can and cannot do.

APPLYING THE QUESTIONS

The simplest and easiest way to identify possible measures for a program (or service or process) is to look for measures that already exist and have been used successfully elsewhere. Such measures are appealing for a couple of reasons:

- Collecting data for an already existing measure is often easier and cheaper than collecting data for a new measure because the methods for collecting data for an existing measure have been developed and tested already.
- The performance of your program on the measure can be compared to the performance of the program for which the measure was originally created as well as other organizations that are using the measure (though this isn't recommended unless the programs are very similar).[7]

6 To see how RBA terms map to logic model terms, see http://raguide.org/3-9-what-is-the-difference-between-4-quadrant-performance-measures-and-logic-model-performance-measures/ and http://raguide.org/framework-crosswalk-analysis/.

As you review an already-existing measure, determine which of the three performance questions it can answer. When you have finished reviewing already-existing measures and compiled a list of the ones that might be useful for measuring the performance of your program, service, or process, determine if any of the three questions cannot be answered by the measures on your list.

Note: One disadvantage of using an already existing performance measure is that it eliminates an opportunity early in an RBA process to engage a program's clients or stakeholders in discussions about what should be measured. Such discussions can lead to more meaningful measures for a program and to greater stakeholder support for the program.

See Appendix B for links to sources of commonly used and accepted performance measures for health and human service programs.

When already existing measures are not available or appropriate for understanding the performance of your program/service/process, create your own measures! Here is one technique that is built around the three types of performance measures:

1. Start by creating a list of the program services or processes you want to understand better. Then, for each service or process on your list, create a measure that will tell you how much of it was accomplished (If, for example, your program delivered training, a measure of how much training it delivered could be the number of people trained).

2. Next, create a measure that will tell you how well each service or process on your list was accomplished (Using training as an example again, a measure of how well your program delivered training might be the percentage of trainees who were highly satisfied with the training or the percentage of trainees who increased their knowledge or skills in the areas in which the program trained them. Alternatively, because program efficiency is also a measure of how well a program performed, you could measure the cost per training, the teacher turnover rate, or the waiting time until the training begins).

3. And finally, create a measure that will tell you if the people who participated in the service(s) or process(es) are any better off afterward. *Note:* "better off" is often defined as an improvement in skills/knowledge, attitudes, behaviors, or circumstances (Using training once more as an example, a measure of the degree to which trainees are better off could be the percentage of them who found and kept a job in the field in which they were trained).

7 For more information on comparing the performance of one program to another — a practice commonly known as external benchmarking — see GovEx's short guide to benchmarking at https://govex.jhu.edu/benchmarking/. While this guide was written primarily with governments in mind, a lot of its advice and resources may be useful to any organization that wants to know how well its programs are doing relative to other programs.

As you create performance measures for your program/service/process, keep the following in mind:

- **Creating performance measures is not a solitary activity!** Ask your program's clients and stakeholders how they would know if a service or process was performed well and how they would know if the clients who received the service(s) were any better off. Also consider asking experts who have studied and written about the service (or similar services). In addition to getting input on performance measures for your program, you may discover a need to clarify your program's objectives or reach more agreement on your program's most important services or processes (tasks you should try to complete early in the RBA process to avoid having to start over if you find out that the measures you have chosen are not suitable for measuring what your program's stakeholders think is important).
- **Consider creating measures that will enable you to (a) assess your progress in solving previously identified issues or problems in a program or (b) answer frequently asked questions about a program.**
- **Some measures can assess more than one type of performance, but they are usually more suitable for assessing one type than another.** For example, a measure might be able to answer, "How well did we do?" as well as "Are our clients any better off?" If you want to measure both types of performance, you may not need to spend time at this stage trying to determine which of these two types of performance the measure is more suitable for. If, on the other hand, you are primarily interested in measuring only one of these types, you can make sure that the measure is more suitable for measuring the type you are interested in by examining the measure in light of your program's primary objectives. For example, if a primary objective of a training program is to increase trainees' knowledge and skills, the percentage of trainees who increased their knowledge or skills is likely to be well suited for measuring whether its clients are any better off. If, on the other hand, a primary objective of the training program is to enable trainees to find and keep a job, then the percentage of trainees who increased their knowledge or skills would be more suitable as a measure of how well the program was performing rather than a measure of its clients are any better off.

TIP

Do not shy away from considering a performance measure that will tell you if your program's clients are better off simply because improving clients' lives is not entirely within your program's control. Making progress on measures like these is rarely within the control of a single program, but that won't stop your program's funders from asking if the program is measuring its performance in this way. Therefore, choose a measure that will tell you if your program's clients are better off, but also make a point of talking about the roles that other organizations can or must play in making progress on that measure. And keep in mind that this type of measure can be very effective in building commitment for your program because it focuses attention on the people that your program is serving and helping.

EVALUATING POSSIBLE MEASURES

After you have compiled a list of possible measures, choose the two to five that are the most important. Any more than five is likely to dilute your focus and increase the amount of effort and resources you will need to spend on data collection.

To help you narrow your options, ask the following questions for each possible measure:

- *Does this measure assess a critical component of my program?*
- *Will this measure give me information about the effect of one or more factors that influence the performance of my program? Can this measure serve as a proxy for other measures?*
- *Is data for this measure available, recent, and accessible? If not, can I develop the tools and methods for collecting the data at a reasonable cost and in a timely manner?*
- *Is this measure sensitive enough to detect small changes in performance? For example, can it detect changes in subgroups of clients?*
- *Will the data for this measure be the same or similar if I collected it from the same source repeatedly? Put another way, does the numerical (or other) value for this measure vary substantially when I collect it repeatedly under the exact same conditions?*
- *Will those who pay attention to my program understand what this measure means?Will the relationship between this measure and what my program aims to accomplish be clear to them? Will this measure answer their questions about my program?*
- *If I had to talk about this program using just a couple of measures, would I include this one?*

None of your possible measures may meet all of these needs, but that is not unusual. **No measure is perfect, so choose the most important ones, and as you use them, be mindful of their deficiencies and the deficiencies in data you get from them.** At the same time, work to address these deficiencies so the quality and value of your data improves over time.

TIP

A good rule of thumb when choosing measures is to pick one or two measures that will answer "Are our clients any better off?" and two to four measures that will answer "How much did we do?" and "How well did we do it?" Note, however, that the actual number and type of measures will vary from program to program depending on the information needs of its staff, clients, and stakeholders, and on the capacity of the organization that runs the program to collect and analyze the data. Another thing to note is that few measures are perfect (especially those that haven't been tried yet). It's quite possible you will modify or change one or more measures as you complete the steps on the following pages. Therefore, do not get bogged down trying to get your measures exactly right at this early stage!

KEY POINTS

STEP 2: IDENTIFYING AND SELECTING PERFORMANCE MEASURES

Three important questions to consider when identifying performance measures are

- How much did we do?
- How well did we do it?
- Is anyone better off?

Answering these three questions is a terrific way to create a comprehensive list of possible performance measures.

These three questions should also be kept in mind when considering performance measures that have been used elsewhere.

Seven additional questions can be answered to select the two to five measures that are most important.

No measure will ever be perfect, and having some data is better than having none. Nevertheless, remembering and acknowledging the limitations of each measure and the data for each measure will go a long way.

This section of this guide focuses on methods for collecting performance data. Although it begins with some tips for finding and using already existing data, most program managers will need to collect new data because already existing data will not be not sufficient or useful. Therefore, most of this section is about collecting new data. It discusses the advantages and disadvantages of the most commonly used data collection methods, points to some excellent online resources and tools for using these methods, and offers some tips for avoiding common causes of inaccuracy in data collection. This section also answers three important, related questions: *(1) How will I know if my program was responsible for improving the health or well-being of the people who received its services? (2) How will I know if my measurements are accurate? (3) How will I know if a difference I have measured is not just an instance of normal variation in measurement?*

FINDING ALREADY COLLECTED DATA

Using already collected performance data, like using already existing performance measures, can save time, effort, and money. If an agency or other organization has data about your client population or service area, you may be able to use it to determine one or more aspects of your program's performance at little or no cost. Data collection is often the costliest part of measuring performance, so using data that another organization has already collected, entered, and verified will save you time and money.[8]

Using already collected data to determine some aspects of your program's performance is advantageous for another reason: it may enable you to compare the performance of your program to other programs for which the same kind of data has been collected.

Here are some suggestions for finding high quality, already existing data that may help you determine your program's performance:

- Review **Appendix C**, which **contains descriptions, links, and major topics for numerous data sets that federal agencies and other organizations make available to the public for free.** A number of these agencies provide data at the county or zip code level (and occasionally even at the community level). Many of these agencies also have staff that will can help you search for and download the data you need.
- **Check out state, city, and local "report cards."** In addition to offering ideas for performance measures you may wish to adopt, the web sites where these report cards reside sometimes make the data behind their measures available to the public.[9]
- If you need economic or labor data (two types of data not well represented in Appendix C), contact your state's census bureau or your local community development corporation.

8 For example, the United Federation of Teachers and the New York City Department of Education conducted a large randomized controlled trial of a teacher incentive program in almost 400 public schools for just $50,000 by using data (student test scores) that participating school districts were collecting already. The only costs for the trial were the costs to prepare, analyze, and report the data (Coalition for Evidence-Based Policy, 2012).

9 For a list of some exemplary report card sites see *http://resultsaccountability.com/about/what-works/*.

TIP

If an agency's data is not available for your client population or service area, ask the agency's staff if they can produce it for you. Although you might not be able to get this data quickly, you can shorten the amount of time agency staff will need to produce it by doing a little homework ahead of time:

- Know specifically what data the agency collects and makes available.
- Know specifically what data you need (refer to your performance measures).
- Be prepared to make the point that you need the data to improve your program, and by improving your program, you are helping the agency achieve its objectives.

COLLECTING NEW DATA

In many (and probably most) cases, you will need to collect new data for your performance measures because already existing data will not be not sufficient or useful. You almost certainly will need to collect new data if you have created your own performance measures.

Your first step in collecting new data is to choose a data collection method. The choice of method is driven, to some extent, by the performance measures you have selected, but you still may have more than one option.

CHOOSING A DATA COLLECTION METHOD

The most common methods for collecting program performance data are

- **Document/Record Review**
- **Focus group**
- **interview (Both Individual and Group)**
- **Observation**
- **Survey**

Not all methods are suitable for collecting data for all kinds of performance measures, but more than one method may be suitable for a given measure. When more than one method appears to be suitable, you can narrow your options based on the amount of time and money you have for data collection, the level of access you have to the people or sites from which you will collect the data, the level of skill required to use the methods you are considering, and the degree of precision and confidence you need in the data you collect.

Table I lists the advantages and disadvantages of each of method.[10] It is important to note, however, that some disadvantages can be overcome by using good data collection practices (discussed in "Improving Data Accuracy" later in this section).

TABLE 1

ADVANTAGES AND DISADVANTAGES OF COMMON DATA COLLECTION METHODS

METHOD	ADVANTAGES	DISADVANTAGES
1 **DOCUMENT/ RECORD REVIEW**	Unobtrusive Usually inexpensive Provides data that is not affected by presence of others Provides data that can be analyzed with statistics (if it is in a standardized format)	Does not allow for follow-up questions Can contain data that is incomplete or inaccurate Can require a lot of time to analyze
2 **FOCUS GROUP**	Enables study of group dynamics Can yield rich data, including new insights Allows for follow-up questions (so works well for complex topics) Provides opportunity to explain or clarify questions Can capture participants' perceptions in their own words	Requires well-trained moderators Requires participants to be in the same location Might not work well for highly sensitive topics Might not permit all topics to be covered (if follow-up questions are asked) Might result in inconsistent data across focus groups (if follow-up questions are asked) Can produce a large amount of information that is difficult to summarize
3 **INTERVIEW** (Both Individual and Group)	Enables interpersonal contact Can yield rich data, including new insights Allows for follow-up questions (so works well for complex topics) Provides opportunity to explain or clarify questions Can incorporate standardized questions that produce data that is consistent across interviews	Requires well-trained Interviewers Might require considerable time and effort Might not permit all topics to be covered (if follow-up questions are asked) Might result in inconsistent data across individuals and groups (if follow-up questions are asked) Can produce a large amount of information that is difficult to summarize Might produce data that is affected by presence of other participants or the interviewer

10 For more information on the advantages and disadvantages of these methods, see the Pell Institute's evaluation toolkit at http://toolkit.pellinstitute.org/evaluation-guide/collect-data/determine-collection-method/ and the Compassion Capital Fund's Capacity Builder's Library, which contains summaries of these and other methods at http://strengtheningnonprofits.org/resources/e-learning/online/datacollection/default.aspx?chp=1.

TABLE 1

ADVANTAGES AND DISADVANTAGES OF COMMON DATA COLLECTION METHODS

METHOD	ADVANTAGES	DISADVANTAGES
3 **INTERVIEW** (Both Individual and Group) — *continued*	Can capture participants' perceptions in their own words	Can produce a large amount of information that is difficult to summarize Might produce data that is affected by presence of other participants or the interviewer
4 **OBSERVATION**	Can capture a wide range of behaviors and events and a wide variety of interactions Provides data that can be analyzed with statistics Enables an understanding of context Can uncover issues that were unknown or unable to be discussed Enables measurement of changes over time	Requires well-trained observers (perhaps even content experts) Might require considerable time and effort Might produce data that is affected by observer bias Might produce data that is affected by presence of observer Might produce data that is atypical (if observations are not repeated)
5 **SURVEY**	Can be conducted quickly and easily (with free online survey tools) Can address a wide variety of topics Can capture participants' perceptions in their own words Can collect data from many people Enables measurement of changes over time Provides data that can be analyzed with statistics	Might produce data that is affected by self-reporting bias Does not usually allow for follow-up questions Does not provide opportunity to explain or clarify questions Might require participants to have access to a computer and an internet connection Might result in insufficient data for analysis (if response rate is low)

In addition to comparing data collection methods based on their advantages and disadvantages, you can use one or more of the following tools to help you choose the best one:

- **A checklist**, such as one or more of the checklists on pp. 24–25 of *Measuring Outcomes*, a guide written by the Compassion Capital Fund's National Resource Center (available at http://strengtheningnonprofits.org/resources/guidebooks/MeasuringOutcomes.pdf).
- **A data collection plan**, such as the two measurement plan templates (one more detailed than the other) created by NHS Education for Scotland (available at https://learn.nes.nhs.scot/3138/quality-improvement-zone/qi-tools/measurement-plan).

ADDITIONAL THOUGHTS

SHOULD I COLLECT DATA FROM A SAMPLE OF INDIVIDUALS OR RECORDS OR ALL OF THEM?

Collecting data from just a sample of individuals or records is useful (and often necessary) when

- The total number of individuals or records is very large — so large that collecting data from all of them would be too expensive or require too much time.
- Some of the individuals or records are not accessible.
- The method of collecting data will impose a burden or other effect on the individuals or records.

With a sufficiently sized sample, you can generalize the results from your sample to the entire population. Put another way, if you want to be confident that the results from your sample are the results you would have gotten if you had collected data from the entire population, the number of individuals or records in your sample needs to be large enough that you can be confident that the results you get from your sample are not simply due to chance[11] (for more on this, see How will I know if a difference I have measured is not just an instance of normal variation in measurement?).

TIP

Determining a sufficient sample size can be as easy as plugging numbers into a formula, but sampling can also be a multi-step and complex process. In such cases, you may save a lot of time by enlisting a statistician to help you.

MORE HELP FOR CHOOSING A DATA COLLECTION METHOD

A number of excellent online resources can help you decide which method is best for collecting data for the measures you have chosen. They can also help you use these methods correctly:

- The Agency for Health Care and Research (AHRQ) has assembled lists containing **basic requirements for conducting surveys, focus groups, and interviews at** https://www.ahrq.gov/professionals/quality-patient-safety/talkingquality/assess/tools.html.
- **Similar lists for conducting surveys and observation**s can be found in the Urban Institute's Key Steps in Outcome Measurement (pp. 16–17), which is available at http://www.urban.org/research/publication/key-steps-outcome-management.
- **A more detailed guide to developing surveys**, written by VPI's Robert Frary, can be found at http://www.indiana.edu/~educy520/sec5982/week_3/questionnaire_development_frary.pdf.

11 This does not mean that data collected from a smaller sample won't provide useful information. Even feedback from just one or two clients can provide valuable information for improving service delivery.

- **More suggestions for conducting observations** are presented in a short paper written by Richard Krueger at the University of Minnesota. It can be found at http://www.betterevaluation.org/en/resources/guide/how_to_use_observation.
- Numerous organizations provide information about, and access to, **already validated surveys that can be easily adapted for other purposes:**
    - AHRQ's Consumer Assessment of Healthcare Providers and Systems (CAHPS) surveys ask consumers and patients to report on and evaluate their experiences with health care. For access and information, see https://www.ahrq.gov/cahps/surveys-guidance/index.html.
    - The Institute for Healthcare Improvement offers two surveys, one for measuring staff satisfaction (available at http://www.ihi.org/resources/Pages/Tools/ProviderandStaffSatisfactionSurvey.aspx) and the other for measuring client/patient satisfaction (available at http://www.ihi.org/resources/Pages/Tools/ShortSurvey.aspx).
    - PerformWell (a collaboration between the Urban Institute, Child Trends, and Social Solutions) has collected over 200 surveys containing measures of education & cognitive development; employment, housing, and self-sufficiency; health and safety (including crime and delinquency, good health habits, mental health, reproductive health, substance use, and victimization); psychological and emotional development; relationships; social and behavioral development; and civic engagement and community involvement. For access and information, see http://www.performwell.org/index.php/find-surveyassessments.
- The American Society for Quality provides **simple instructions for using check sheets to carry out observational studies** at http://asq.org/learn-about-quality/data-collection-analysis-tools/overview/check-sheet.html (this site also provides a spreadsheet tool for entering check sheet data and automatically generating charts displaying this data).
- The RAND Corp's **manual for conducting semi-structured interviews and focus groups** is at https://www.rand.org/pubs/technical_reports/TR718.html.
- The Institute for Healthcare Improvement has created an "Improvement Tracker," which is **an online tool for entering, tracking, graphing, and reporting data** for a host of pre-selected measures — but can also be used for one's own measures. It is available at http://app.ihi.org/Workspace/tracker/ (free, but registration is required).

DETERMINING EFFECTIVENESS

In many cases, collecting performance data just for the program, service, or process you are studying will be enough. For example, if you want to know if productivity has reached some target or quality has stayed above some threshold, you can simply collect data for the program, service, or process over time and analyze it for meaningful differences between

time points. Monitoring performance in this way can yield valuable information about how a program, service, or process is operating, the circumstances in which it works best, and its acceptability and appropriateness for its users.

However, if you want to know with any certainty that an improvement in the performance of a program/service/process was the result of some modification you made in it (in other words, that what you modified in it was effective), you may need to compare the performance of the modified program/service/process to the performance of an unmodified, but otherwise very similar, program/service/process. Without this kind of comparison, you may never know if performance might have improved not only in the modified program/service/process but also in the unmodified one — a result that would strongly suggest that something other than the modification was responsible for the improvement. Put a different way, if performance improved in the modified program/service/process but not in the unmodified one, you can conclude that the modification was responsible for the improvement. If, on the other hand, performance improved in both programs/services/processes equally, it is likely that something other than the modification was responsible for the improvement.

A challenge in making such comparisons is selecting a program, service, or process that is very similar to the one that you are modifying. If the modified program/service/process differs from the unmodified one in some way other than the modification, this other difference — rather than the modification — might be responsible for any improvement in performance. However, you can minimize the presence of such differences in a number of ways. For example, you can run the program/service/process with and without the modification and then compare the two. Or, if the program/service/process is something that you run often, you can modify it randomly (for example, by flipping a coin to determine when you will and won't modify it). Randomizing the modification (which, in this case, is analogous to randomizing an intervention) will increase confidence that the only significant difference between the programs/services/processes that will be modified and the ones that won't is the modification in the former.[12]

When a comparison to a very similar program/service/process is not possible, you may still be able to establish that a modification is responsible for improving performance — if you can rule out other plausible explanations for the improvement. But even when you cannot rule out all other plausible explanations, an otherwise well-designed study can still provide you with preliminary evidence that a modification is responsible for an improvement in performance.

12 Randomizing the modification will not, however, eliminate the possibility of a difference between the individuals or items serving as the measure of performance for the modified programs/services/processes and the individuals or items serving as the measure of performance for the unmodified ones. This kind of difference can be present even when the only difference between the modified and unmodified programs/services/processes is the modification itself. You can minimize such differences by placing the individuals/items into the modified and unmodified programs/services/processes randomly (e.g., by flipping a coin). This will increase confidence that (a) the individuals/items placed in each program/service/process are highly similar and (b) any improvement in the performance of the modified program/service/process is the result of the modification rather than the result of a difference in the individuals/items placed in each program/service/process. When random placement is not possible, you can still attribute an improvement in the performance of a program/service/process to a modification that you made in it if you can rule out any significant differences between the individuals/items in it and the unmodified version.

ADDITIONAL THOUGHTS

HOW WILL I KNOW IF MY PROGRAM WAS RESPONSIBLE FOR IMPROVING THE HEALTH OR WELL-BEING OF ITS CLIENTS?

Measuring improvements in the health or well-being of your program's clients is one thing; determining if your program was responsible for these improvements is another thing altogether. It is not hard to find examples of programs that appeared at first to improve their clients' health or well-being but in later studies showed that was not the case.[13]

The surest way to establish that a program is at least partly responsible for improving the health or well-being of its clients is to randomly assign individuals either to a group that will participate in the program or a comparison group that will not. Randomly assigning individuals to these "treatment" and "control" groups ensures that the two groups are alike in all respects apart from receiving the program's services. When two groups differ only in the services they receive, any difference in health or well-being between the two groups can be attributed directly to the services (assuming the number of people in each group is large enough. For more information, see How will I know if a difference I have measured is not just an instance of normal variation in measurement?).

In practice, however, a randomized controlled trial (or at least a traditional one) is often not possible or appropriate for one or more reasons: (a) it would be too expensive;[14] (b) it would be unethical (for example, if a service that was already proven to be effective was given to one group but not the other); (c) the individuals in one of the groups could drop out at a high rate, potentially biasing the trial results;[15] or (d) the individuals in the comparison group might have access to the services being studied. Although many program managers are quite familiar with these obstacles to carrying out a conventional randomized controlled trial, **fewer program managers are aware of randomized controlled trial variations that can overcome some or all of these obstacles without compromising their ability to attribute a difference in health or well-being to a program.** One example is the crossover trial, in which all enrollees in the trial receive a program's services, but they receive

13 Some notable examples are the Drug Abuse Resistance Education (D.A.R.E.) program (Perry et al., 2003), the Scared Straight program (Petrosino, Turpin-Petrosino, & Finckenauer, 2000), and Upward Bound (U.S. Department of Education, 2004).

14 While randomized controlled trials can be expensive, they aren't always so. Before ruling out a randomized controlled trial because of its cost, consider the following opportunities to lower costs:

- Instead of paying for the collection of new data (often the biggest cost in a trial), use already existing data (e.g., analyze student test scores that are being collected already for other purposes).
- Instead of paying for all operating costs of a trial (e.g., information technology, supplies, and personnel), use resources that are funded already as part of normal operations (e.g., salaries).
- If a program cannot be offered immediately to every individual or site eligible for it, use randomization to select which individuals or sites receive it first (for example, the Metropolitan Police Department of the District of Columbia carried out a relatively low-cost randomized controlled trial of the effectiveness of police body cameras by randomizing the distribution of the cameras by district - something that could be accomplished easily and cheaply because District officials had already made plans to stagger the distribution of the cameras by district for logistical reasons (Yokum, Ravishankar, & Coppock, 2017).
- Statistical analysis of trial data, though not necessarily expensive, does require some expertise that may not be available within your organization. RCT-Yes, a free tool for estimating and reporting the effect of a program when it is evaluated in a randomized controlled trial or other comparison group study, is available at https://www.rct-yes.com/. A somewhat similar tool that guides users through an evaluation of educational technology (RCE Coach) is available at https://www.edtechrce.org.

15 For example, individuals participating in a trial of an experimental group therapy program sometimes drop out of the program at a high rate when they realize that some members of their group are making the behavioral changes that the program is encouraging, but they are not.

each service in a different sequence (with the order determined randomly). Another example is the cluster trial, in which geographically separated locations (e.g., communities, clinics, schools) rather than individuals are randomly assigned to either a program's services or standard services. Other randomized trial variations capable of overcoming the obstacles to carrying out a conventional randomized trial include the encouragement trial and the delayed start trial.[16]

When random assignment is not possible, a study using an appropriate comparison group could be a sound alternative. In this type of study, a group of individuals who received a program's services is compared to a similar group of individuals who did not (such as a group of individuals in a neighboring community). However, this type of study can sometimes misjudge the effect of the program if the two groups it is comparing differ markedly in some demographic characteristic, ability or skill, or behavior (e.g., motivation) that influences improvement. In such cases, any improvement in the group receiving services may be due, at least in part, to this difference rather than the services. However, when the two groups are very similar with respect to these influences, the chance that a non-randomized study will produce valid results increases substantially.[17] Furthermore, **highly similar comparison groups aren't always difficult to find and recruit.** Individuals on a program's wait list and individuals that are just below the threshold for enrollment in a program's services are just two examples of highly similar comparison groups that may be close at hand.

When no comparison group can be identified, the next best option may be to simply measure health or well-being in a group of individuals before and after they receive the services under study. However, because this type of study lacks a comparison group, there is no way to know if health or well-being might have improved in individuals who *did* not receive the program's services — a result that would strongly suggest that something other than the services was responsible for the improvement. Without a comparison group, there is also no way to know if health or well-being would have worsened if individuals did not receive the services (as might be the case if the services were effective, but other changes in the environment of the individuals who received those services worsened their health or well-being more than the services improved it). Nevertheless, "before/after" studies can still provide preliminary evidence for the effectiveness of a program's services, especially if (a) the difference in health or well-being after individuals receive services is substantial and detected soon after services are delivered and (b) other plausible explanations for the difference can be ruled out.

IMPROVING DATA ACCURACY

"Accurate" data are data that reflect the true value of whatever has been measured. **Data that are not accurate can be hard to detect, but fortunately, they can be avoided (or at least corrected for). Two common causes of inaccurate data are flaws in the *instruments* that are used to collect and process the data and flaws in the *procedures* that are used to collect and process the data.** In interviews, for example, the interviewer may unconsciously influence responses, record the data incorrectly, or fail to adhere to a script.

16 For a more thorough (but still relatively brief) review of the most common variants of the randomized controlled trial (and the various types of comparison group studies), see https://www.npcnow.org/publication/making-informed-decisions-assessing-strengths-and-weaknesses-study-designs-and-analytic.

17 Several literature reviews have identified additional factors that increase the likelihood that a non-randomized study will produce valid results, including (a) collection of data in the same way for both groups and (b) adjustment for minor differences in group characteristics using statistical methods. For more information, see http://coalition4evidence.org/wp-content/uploads/2014/01/Validity-of-comparison-group-designs-updated-January-2014.pdf.

In surveys, questions may be ambiguous, instructions may be confusing, or responses may be incorrectly coded or tabulated. Incorrect administrative records and biases in sample selection are additional examples of flaws in data collection.

A key characteristic of these types of flaws (otherwise known as "systematic errors") is that they produce data that is consistently above or below the correct value. This data can differ significantly from data collected at another time or from another group, program, service, or process. When this happens, the difference in the two sets of data can be mistakenly attributed to a difference between the two groups, programs, services, or processes or to something that changed between the time points (such as the delivery of services to a program's clients) when the difference in data was actually the result of a flaw in a data collection instrument or procedure.

The good news is that you can avoid these mistakes by using good data collection practices:

- **Before Data Collection**
  - Use multiple measures to answer each performance question (when possible)
  - Pilot test any instrument(s) for collecting data (e.g., ask colleagues to review or complete a survey you have designed)
  - Thoroughly train any individuals who will collect, record, and/or transfer data
  - Randomly choose any samples that will be taken
  - Randomly assign individuals/records to any treatment and control groups that will be used
- **After Data Collection (During Data Entry)**
  - Use the data validation features in spreadsheet and database programs to ensure uniform entry of data (e.g., by restricting fields or using drop-down lists containing all possible values for a field)
  - Enter data twice and then compare the two sets of entered data
  - Compare a randomly selected subset of entered data to the originally collected data
  - Sort entered data to find missing or outlying data

ADDITIONAL THOUGHTS

HOW WILL I KNOW IF A DIFFERENCE I HAVE MEASURED IS NOT JUST AN INSTANCE OF NORMAL VARIATION IN MEASUREMENT?

Measurement is almost always imperfect, even when no flaws exist in the data collection instruments or procedures. Imperfect measurement has many causes: imprecision in data collection instruments, inconsistency among those who use the instruments and record the data, inconsistency in the actions of those from whom the data is collected, and other factors – most of which are unexpected and unpredictable (and therefore uncontrollable). Because of these factors, you are likely to see differences in the measurements you take when your repeat them, even when you repeat them using the same instrument or take them from the same individual under the same conditions. This is often called "normal" variation, and it is usually completely random in both the positive and negative direction around the correct measurement.[18] Consequently, it is cancelled out when you take measurements numerous times and calculate the average[19] value of them. In other words, **the average of a set of repeated measurements is a good estimate of the correct measurement.**

The average is a useful statistic for other reasons too:

- **The amount of variation in a set of measurements — known as the standard deviation — can be determined using just the value of the average and the values of the individual measurements that make up that average.**
- **The probability that the difference between the averages for two groups, programs, or time points is not an instance of normal variation in measurement can (with some exceptions) be calculated using the student's t-test, a formula that requires just the averages, standard deviations, and sizes of the groups, programs, or time points measured.**

A related imperfection in measurement is specific to measurements taken from a sample of individuals in a population. Because of the diversity/variation in any given population, measurements taken from a sample of individuals in a population will differ to some extent from measurements taken from all individuals in the population, which means that the average of the measurements taken from the sample is likely to differ to some extent from the average of the measurements taken from the entire population. However, the larger the size of the sample, the closer the average of measurements for that sample will come to the average of measurements for the entire population. In addition, the range of measurements in which the average measurement for the entire population is likely to lie (termed the *confidence interval* for the average) can be calculated using the value of the average, the individual values that make up the average, and the standard deviation for the average.

18 Note that differences in measurement resulting from these causes are distinct from differences that result from flaws in data collection instruments or procedures, where the differences in measurement are consistently above or below the correct measurement (see Improving Data Accuracy).

19 Technically, this statistic is the mean rather than the average because the mode and median are averages too.

KEY POINTS
STEP 3: COLLECTING DATA FOR PERFORMANCE MEASURES

Using already-collected data, like using already-existing measures, can save time, money, and effort, but it is rarely trouble-free, and in many cases, you will want (or need) to collect new data.

Choosing a data collection method is driven, to some extent, by the performance measures you have selected, but you may still have more than one option. Table I summarizes the advantages and disadvantages of the most common data collection methods. When in doubt, use the checklist, data collection planning tool, or method-specific resources described in this section to choose a method.

Collecting performance data for a program, service, or process over time and analyzing it for meaningful differences between time points can yield valuable information about how a program, service, or process is operating, the circumstances in which it works best, and its acceptability and appropriateness for its users.

The surest way to know that an improvement in the performance of a program/service/process was the result of a modification in it is to compare the performance of the modified program/service/process to the performance of an unmodified, but otherwise very similar, program/service/process.

When a comparison to a very similar program/service/process is not possible, it is still possible to establish that a modification is responsible for improving performance by ruling out other plausible explanations for the improvement. But even when all other plausible explanations cannot be ruled out, an otherwise well-designed study can still provide preliminary evidence that a modification is responsible for improving performance.

The surest way to establish that a program is at least partly responsible for improving the health or well-being of its clients is to randomly assign individuals either to a group that will participate in the program or a comparison group that will not. Numerous randomized trial variations can overcome some or all of the common obstacles to carrying out conventional randomized trials.

When random assignment is not possible, a study using an appropriate comparison group may be a sound alternative, especially since appropriate comparison groups aren't always difficult to find and recruit.

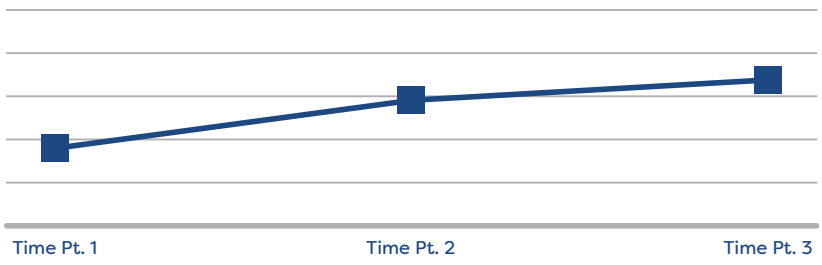
Good data collection practices and a few basic statistical formulas can help prevent the mistake of assuming that a change in performance measurements means that performance changed when it was actually just normal variation in measurement or a flaw in data collection.

After completing the first three steps of the RBA process, you are likely to have some data that contain a lot of interesting and useful information regarding the performance of your program. But how do you make sense of it? How do you organize it, scrutinize it, and use it to understand your program's performance? This section will show you ways to do all of these things.

ANALYZING QUANTITATIVE DATA

Analysis of quantitative data (that is, data expressed as numbers) almost always begins by compiling the data in a table, or better yet, a graph or figure. Why? Because **a visual display of data can improve the ability to spot relationships and patterns.** A relatively simple case is shown in Figure 2, which plots the average value of a set of measurements collected at three different time points. A graph like this could show, for example, the average value of some measure of the productivity, quality, or effectiveness of a program/service/process over time, or the average value of some measure of the health or well-being of a program's clients over time. As noted in the subsection entitled "Determining Effectiveness, graphing data for a single program, service, or process can yield valuable information about how it is operating, the circumstances in which it works best, and its acceptability and appropriateness for its users.

FIGURE 2

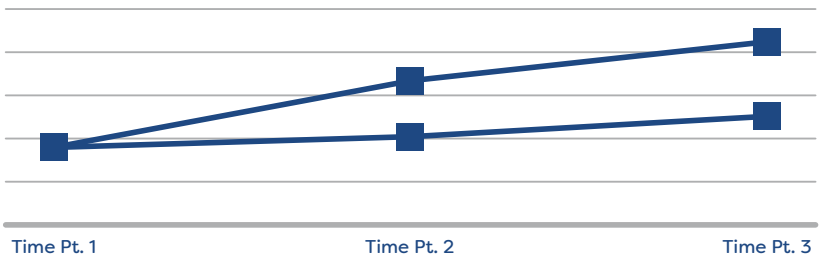


However, in that same section, it was also noted, that it may be necessary to graph and compare the data for two programs/services/processes if the purpose of collecting the data is to determine whether a modification to a program, service, or process *is responsible for* improving performance (in other words, whether the modification is *effective*). Similarly, it may be necessary to graph and compare the data for two groups of individuals if the purpose is to determine if a program, service, or process is responsible for (in other words, *effective* in) improving the health or well-being of the individuals who participated in it.

Figure 3 shows such a comparison. Let's say that the figure shows the values for some measure of health or well-being in two groups of individuals, but only one of these groups received certain services. If the individuals in each group are very similar (because, for example, each individual was randomly placed into one group or the other), you can be confident that the services were responsible for changing their health or well-being. You also can be confident that the services were responsible for the *difference* in health or well-being

Now let's say that Figure 3 shows the values for some measure of performance for two programs, services, or processes, but a modification was made in only one of them. If the only significant difference between the two programs/services/processes is this modification, you can be confident that the modification was responsible for the improvement in performance *and* it was also responsible for the *difference* in performance *between* the modified program and the unmodified one.[20]

FIGURE 3



TIP

Try graphing the data by subgroups to see if differences exist between them. For more information on sorting (stratifying) data into subgroups, see http://asq.org/learn-about-quality/data-collection-analysis-tools/overview/stratification.html.

SOME ALTERNATIVE CHART TYPES AND THEIR USES	
FREQUENCY HISTOGRAM/ BAR CHART	Shows the frequency with which different measurement values (e.g., times, weights, temperatures) occur in a sample. By doing so, it shows the distribution of values in the sample. For more information, see http://asq.org/learn-about-quality/data-collection-analysis-tools/overview/histogram.html.
CONTROL CHART	Shows the frequency with which different measurement values (e.g., times, weights, temperatures) occur but also displays these values over time and includes lines that mark the upper and lower limits of what has been determined to be normal variation (usually in a process that is being studied). Data points lying outside these lines are considered non-normal variation. For more information, see http://www.syque.com/quality_tools/toolbook/Control/how.htm.
BOX AND WHISKER PLOT	Uses the middle value of a set of data (the median) and the values of quartiles of that data to show the amount of variation in the data. For more information, see http://asq.org/learn-about-quality/data-collection-analysis-tools/overview/box-whisker-plot.html

20 Although comparison studies can rule out many plausible explanations for an improvement in the performance of a program, service, process, they cannot rule out two things that may produce what appears to be an improvement but really isn't:

- A flaw in a data collection instrument or procedure
- Normal/naturally occurring variation in measurement

As discussed in the subsection entitled "Improving Data Accuracy", however, good data collection practices can minimize the possibility of a flaw in a data collection instrument or procedure, and statistical formulas (e.g., the student's t-test) can be used to calculate the probability that normal/naturally occurring variation, rather than the program, service, or process, is responsible for a change in performance measurements.

ORGANIZING AND ANALYZING QUALITATIVE DATA

All data collection methods described in this guide can provide qualitative data (that is, data in words rather than numbers). **Qualitative data can provide a rich amount of information, but it can also be time-consuming to organize, especially if there is a lot of it** (as is sometimes the case with interview transcripts, for example). Some types of qualitative data can be converted into numerical data that is easier to analyze, but other types cannot (or should not) be converted. **Here is a general approach for organizing and analyzing qualitative data thoroughly *and* efficiently:**

1 Read the text (transcripts, surveys responses, etc.) several times, writing down any impressions that come to mind (these impressions may save you time later).

2 Based on your reading and your rationale for collecting the data, identify the most important questions you want to answer (that is, choose a small number of critical questions).

3 Create a set of categories that reflect possible answers to the questions you have identified in step 2 above. Then read the text again, this time focusing on these answers and placing them in the appropriate categories (alternatively, you can create categories as you read). Try to eliminate categories that overlap by adding, combining, and splitting categories as needed (but still feeling free to place answers in more than one category when you need to).

4 Count the number of items in each category. Although these counts are not suitable for statistical analysis, they will give you a rough estimate of the importance of each category.

5 Look for patterns in the answers you have placed in each category. Also look for similarities and differences in answers between categories. Take note of answers that consistently occur together. While you should be cautious in concluding that co-occurring items share a cause and effect relationship, co-occurrences can still help explain why something occurs.

When analyzing qualitative data, keep the following pitfalls in mind:

- In most cases, conclusions drawn from qualitative data cannot be generalized to other conditions, settings, or individuals because the data is usually derived from very small samples, and because the analysis of qualitative data is, to some degree, always subjective. The power of qualitative data lies in its richness; it provides an opportunity to understand the real meaning of a comment, response, or behavior — something that is frequently impossible when analyzing quantitative data. In addition, many methods for collecting qualitative data provide opportunities to ask follow-up questions that can yield new areas of study as well as clarifying information.
- Quotes from interviews and focus groups can serve as vivid examples of typical responses or how responses differ, but they are often misused. Avoid editing quotes to better support

conclusions, and always request permission to use a quote, even if the quote will be anonymous (a reader still might be able to guess the identity of the individual who was interviewed).

Whole textbooks have been written about qualitative data analysis, so this short guide hardly scratches the surface of it. For more information about qualitative data analysis, check out these resources:

- The Pell Institute's Evaluation Toolkit describes **a six-step process** for analyzing qualitative data at http://toolkit.pellinstitute.org/evaluation-guide/analyze/analyze-qualitative-data/.
- The University of Wisconsin Cooperative Extension's very user-friendly guide for analyzing qualitative data provides **detailed instructions, with numerous examples, for categorizing qualitative data and using categorization to identify themes and patterns in data.** It also provides tips for improving the process and presents a number of potential pitfalls. While this guide predates some software tools that make qualitative analysis easier, it is still widely used. You can find it here: https://learningstore.uwex.edu/assets/pdfs/g3658-12.pdf.
- The Happy City Program in the United Kingdom has produced a **guide covering the collection and analysis of qualitative data on well-being specifically**. You can find it here: http://www.happycity.org.uk/wp-content/uploads/2016/10/Happy-City-Wellbeing-Measurement-A-Guide-to-Qualitative-Data-Collection.pdf
- SocialCops (a private company offering a data intelligence platform) hosts a blog where a nice **guide for converting qualitative data into numerical data has been posted.** You can find it here: https://blog.socialcops.com/academy/resources/guide-quantifying-qualitative-data/.

ASSESSING PAST PERFORMANCE

Previous sections of this guide have described different types of performance and how to go about measuring them. Regardless of the kind of data you have collected and the type of performance you have measured, **the questions you should answer about the data are the same:**

- *What does this data tell us about the performance of the program (or service or process)?*
- *Is performance better or worse than we expected?*
- *Are these the right/best performance measures?*
- *What more would we like to know?*

TIP

While getting more or better data is sometimes essential, at other times it is not, and instead it is "deciding not to decide." If you have used good data collection practices, your conclusions based on your data are sound, and you have enough information in hand to act, do so!

FORECASTING FUTURE PERFORMANCE

Once you have a better grasp of past performance, forecast what you expect performance to look like in the future.[21] When doing this, make sure you adjust for

- any anticipated or planned changes in how the program/service/process is implemented
- any external factors that might affect performance
- any lag in performance due to the amount of time that is needed for the program/service/process to affect what you have measured

Keep a record of these factors (you will use them later).

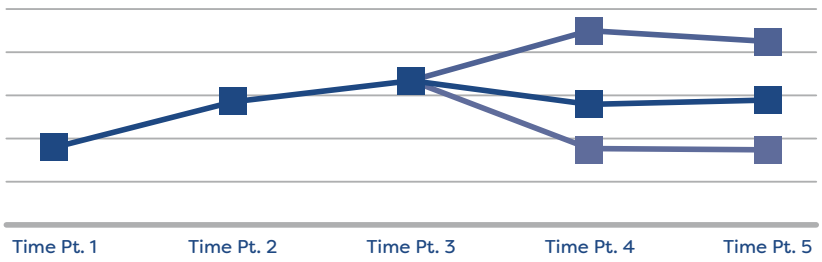
TIP

To get the most believable and accurate forecast, talk to people who might know more about the external factors that might affect future performance. Also talk to people who might know more about the time required for the program/service/process to have a measurable effect.

21 For in-depth guidance on using past performance data and other information to accurately predict future performance, see GovEx's guide to setting performance targets (especially its information on practicing accuracy). It is available at https://centerforgov.gitbooks.io/setting-performance-targets-getting-started-guide/content/.

If more than one forecast is reasonable, add it to your graph. Figure 4, for example, shows three possible forecasts (time points 4 and 5) for the performance measure shown in Figure 2.

FIGURE 4



Once you have forecasted performance, ask yourself (and others who know the program/service/process well) these questions:

- *Is forecasted performance acceptable?*
- *Could forecasted performance be better?*

FIGURING OUT THE STORY BEHIND PERFORMANCE

Many people who use RBA cite its emphasis on telling "the story behind performance" as one of its most appealing features. To tell that story, however, you need to know the root causes of performance and the forces that drive performance. Unfortunately, these causes and forces aren't always apparent, and **it is quite common for program managers, staff, and others to mistake superficial causes for root causes, or to tackle superficial causes and forces without trying to identify the underlying ones.** Such efforts are unlikely to be the most effective ways to improve performance.

Classic problem-solving and quality improvement tools can help you avoid such mistakes by uncovering the driving forces and root causes of performance. They may also bring to light any number of common biases and assumptions that can thwart good decision making. Here are brief descriptions of some of the best-known of these tools, with links to instructions for using them.:[22]

- **The Cause and Effect (aka "Fishbone") Diagram** is a tool for brainstorming all the potential causes of a problem in performance and sorting them into useful categories. For more information, see http://asq.org/learn-about-quality/cause-analysis-tools/overview/fishbone.html.

22 A tutorial explaining several of these tools is available from the National Quality Center at http://nationalqualitycente2222r.org/resources/nqc-quality-academy-useful-quality-improvement-tools/.

- **The Five Whys Technique** focuses on identifying the root causes of a performance problem and is notable for its simplicity. However, it has limited ability to explore more than one idea that might lead to a root cause. For more information, see https://www.mindtools.com/pages/article/newTMC_5W.htm.

- **The Process Flowchart** is used primarily to identify all steps in a process. It is usually used at the beginning of a process improvement effort. For more information, see http://asq.org/learn-about-quality/process-analysis-tools/overview/flowchart.html.

- **The Relations Diagram** is a more advanced tool for identifying the potential causes of some effect and the relationships between these causes. It is often used after other methods for identifying causes have been attempted, or when an issue is particularly complex. For more information, see http://asq.org/learn-about-quality/new-management-planning-tools/overview/relations-diagram.html and http://asq.org/healthcare-use/why-quality/relations-diagram.html

- **A Force Field Analysis** identifies potential causes as well, but it goes a step further by asking users to estimate the importance of each cause. Comparing these "weighted" causes head to head can help users decide where to focus their improvement efforts. For more information, see http://www.phf.org/resourcestools/Pages/Diagnostic_Force_Field_Analysis.aspx.

- **The Consensus Decision Making Matrix** is useful when a group is struggling to come to agreement on all significant causes for a program's performance. Its matrix provides a structure for all members of the group to assess each cause as well as the components of each cause. For more information, see http://www.phf.org/resourcestools/Documents/Consensus_Decision_Making_Matrix.pdf.

These tools may yield better information when individuals work together on them rather than separately. Public Profit's free downloadable guide entitled *Dabbling in the Data* (http://www. publicprofit.net/Dabbling-In-The-Data) provides step-by-step instructions and materials for leading a team or group through several of the tools listed here as well as numerous others, including some that are useful for organizing and analyzing data.

Note: An important consideration in any search for the story behind performance is whether expectations for performance were overly optimistic. Some levels of performance, such as those linked to ambitious program objectives, may require more time to be reached than originally estimated. In these cases, adjusting or modifying a program too soon (or too often) may prevent the program from ever doing all the things that are necessary to reach the level of performance expected. More evidence may need to be gathered to be confident that a change in a program is warranted (for example, data showing at least some improvement in a program's performance over time might indicate that making a major change in the program could be premature even if that performance wasn't as good as expected).

KEY POINTS
STEP 4: ANALYZING AND INTERPRETING PERFORMANCE DATA

Qualitative data from interviews, focus groups, and surveys can provide a rich amount of information, but it can also be time-consuming to organize and analyze. A general approach to analyzing qualitative data thoroughly and efficiently in five steps is provided in this section along with descriptions and links to more comprehensive sources of information.

Regardless of the type of data collected, the questions about the data are largely the same:

- What does this data tell us about the performance of the program (or service or process)?
- Is performance better or worse than we expected?
- Are these the right/best performance measures?
- What more would we like to know?

Forecast future performance by adjusting for

- any anticipated or planned changes in how the program/service/process is implemented
- any external factors that might affect performance
- any lag in performance due to the amount of time that is needed for the program/service/process to affect what you have measured

Then answer these questions:

- Is forecasted performance acceptable?
- Could forecasted performance be better?

Many people who use RBA cite its emphasis on telling "the story behind performance" as one of its most appealing features. To tell that story, however, you need to know the root causes of performance and the forces that drive performance.

The root causes and driving forces behind performance aren't always apparent, and it is quite common for program managers, staff, and others to mistake superficial causes for root causes, or to tackle superficial causes and forces without trying to identify the underlying ones. Such efforts are unlikely to be the most effective ways to improve performance. Classic problem-solving and quality improvement tools can help you avoid such mistakes by uncovering the driving forces and root causes of performance. They may also bring to light any number of common biases and assumptions that can thwart good decision making.

Once you have figured out the story behind the past (or expected future) performance of your program, service, or process, you are ready to identify opportunities and options for improving that performance. You can find these opportunities and options both inside and outside your organization.

FINDING MORE EFFECTIVE PROGRAMS AND SERVICES

Finding more effective programs and services has become much easier over the past decade because of the availability of numerous "what works" clearinghouses online. Here are descriptions of, and links to, some of the best of these clearinghouses:

- Child Trends has created **What Works** (https://www.childtrends.org/what-works/), a searchable registry of over 700 programs that have been evaluated in at least one randomized controlled trial that showed an improvement in education, life skills, or social/emotional, mental, physical, behavioral, or reproductive health in children or youth.
- The Pew-MacArthur Results First initiative has created **The Results First Database** (http://www.pewtrusts.org/en/multimedia/data-visualizations/2015/results-first-clearinghouse-database), a one-stop source of information on the effectiveness of interventions that were rated by at least one of seven national research clearinghouses (Blueprints for Healthy Youth Development, California Evidence-Based Clearinghouse for Child Welfare, Coalition for Evidence-Based Policy, CrimeSolutions.gov, National Registry of Evidence-Based Programs and Practice, Promising Practices Network, What Works Clearinghouse, and the What Works in Reentry Clearinghouse). Designed with policymakers in mind, the Results First Database has reconciled the different rating systems and vocabularies of these clearinghouses and provided their data in a clear, accessible format.
- **Social Programs that Work** (http://evidencebasedprograms.org/), sponsored by the Coalition for Evidence-Based Policy, provides information about programs that have demonstrated — via randomized controlled trials — sizable, sustained effects. Although the number of programs meeting this standard is less than 30,[23] the site covers a wide range of areas: prenatal/early childhood, K-12 education, postsecondary education, teen pregnancy prevention, crime/violence prevention, housing/homelessness, employment and welfare, substance abuse prevention/treatment, obesity and disease prevention, mental health, and health care financing/delivery.
- The Office of Disease Prevention and Health Promotion at the U.S. Department of Health and Human Services maintains **a database of evidence-based interventions and other resources** (https://www.healthypeople.gov/2020/tools-resources/Evidence-Based-Resources) that can enable progress toward Healthy People 2020 targets. This database contains over 400 entries, each one with a rating that corresponds to the strength of evidence supporting its effectiveness, feasibility, reach, sustainability, and transferability.

23 As of June 2018

- The Centers for Disease Control and Prevention (CDC) maintains **the Community Guide** (https://www.thecommunityguide.org/), a repository for the findings of the Community Preventive Services Task Force (CPSTF). The CPSTF reviews intervention approaches across a wide range of health topics to determine how well they work, what they might cost to implement, and what gaps in evidence for their effectiveness have not been filled yet.

In addition to the rigorously evaluated programs, interventions, and services that are the focus of these clearinghouses, a much larger number of promising programs, interventions, and services are supported by some objective data on their effectiveness, but they have not been evaluated in a randomized controlled trial or a trial that used a highly similar comparison group. While these programs, interventions, and services are well worth consideration, it is important to closely examine the strength of the evidence supporting their effectiveness first (Ng & de Colombani, 2015).[24]

You should document the following information for any programs, interventions, or services that you are considering:

- the context in which the program/intervention was shown to be effective
- the length of time that the program/intervention was implemented
- the tools, processes, and systems that supported implementation of the program/intervention

Collecting this information will enable you to determine the extent to which you can replicate the effectiveness of the program, intervention, or service in your location and with the resources you have.[25]

FINDING OTHER WAYS TO IMPROVE PERFORMANCE

Improving effectiveness is important, but a program, service, or process can be effective and still fall short. Process inefficiencies and poor quality control can do great harm to an otherwise sound activity, but the "what works" clearinghouses are unlikely to be of much help in such circumstances. **The best ideas for solving these types of performance problems are likely to lie within your organization, among your organization's clients, or in other organizations that administer activities similar to the one you wish to improve.**[26] To uncover ideas from these sources, share and discuss the story behind performance with them. Ask them for their ideas for improving performance. A casual conversation may be sufficient, but if you structure your discussion with

24 A number of publications do an excellent job of explaining various types of evidence and how to assess the strength of evidence supporting a program's effectiveness. One of them is the CDC's "Understanding Evidence Part 1: Best Available Research Evidence. A Guide to the Continuum of Evidence of Effectiveness" (https://www.cdc.gov/violenceprevention/pdf/understanding_evidence-a.pdf). Two others are Mathematica's *Understanding Types of Evidence: A Guide for Educators* (https://www.mathematica-mpr.com/our-publications-and-findings/publications/understanding-types-of-evidence-a-guide-for-educators) and The North Carolina General Assembly's *Ten Questions to Better Pilot Programs* (http://www.ncleg.net/fiscalresearch/fiscal_briefs/Fiscal_Briefs_PDFs/Getting_More_From_Pilot_Programs_Fiscal_Brief_FINALweb.pdf).

25 For a more in-depth discussion on documenting information about programs of interest, see Identifying and Promoting Effective Practices, a publication available from the Compassion Capital Fund's National Resource Center at http://strengtheningnonprofits.org/resources/guidebooks/Identifying%20and%20Promoting%20Effective%20Practices.pdf)

26 Another source of ideas for improving performance is the web site for the RE-AIM Framework, which focuses on five areas driving a program's public health impact: Reach, Effectiveness, Adoption, Implementation, and Maintenance. Links to ideas for improving performance in each of these areas can be found at http://www.re-aim.org/about/what-is-re-aim/.

them using one or more of the following problem-solving techniques, you are likely to get more (and sometimes better) ideas:

- **The Nominal Group Technique** is a method for brainstorming that enables contributions from every participant — not just the most vocal and quick-thinking ones. For more information, see http://asq.org/learn-about-quality/idea-creation-tools/overview/nominal-group.html.[27]
- **The SCAMPER Technique** rolls seven different thinking approaches into one (SCAMPER stands for Substitute, Combine, Adapt, Modify, Put to another use, Eliminate and Reverse). For more information, see http://www.designorate.com/a-guide-to-the-scamper-technique-for-creative-thinking/.
- **A Community Stakeholder Services Map** can assist staff who manage a health program identify opportunities for improving the program by leveraging changes in local transportation, parking, public safety, and accessibility. For more information, see http://www.phf.org/resourcestools/Pages/Community_Stakeholder_Services_Map.aspx.

TIP

When you ask colleagues and others for ideas, encourage them to suggest ideas that you can implement at low or no cost.

TIP

Use extra care when sharing data about program performance with the staff who are responsible for implementing the program. Make the point that (a) the data gives them information that will help them be more successful and (b) knowing how they are doing is valuable. Treat the data as an opportunity for learning rather than judgment. And don't neglect to share positive findings in addition to negative ones!

27 For instructions on using the more traditional style of brainstorming, see http://asq.org/learn-about-quality/idea-creation-tools/overview/brainstorm.html.

KEY POINTS
STEP 5: FIGURING OUT WHAT MIGHT WORK TO IMPROVE PERFORMANCE

Finding more effective programs and services has become much easier over the past decade because of the availability of numerous online "what works" clearinghouses.

Documenting the following information for any programs or services that appear promising can improve the ability to replicate their effectiveness in a different location or context with the resources that are available:

- the context in which the program/intervention was shown to be effective
- the length of time that the program/intervention was implemented
- the tools, processes, and systems that supported implementation of the program/intervention

The best ideas for improving the productivity, efficiency, or quality of a program, service, or process are likely to lie within your organization, among your organization's clients, or in other organizations that operate programs, services, or processes that are similar to the one you wish to improve. Using structured problem-solving techniques will increase the odds of getting more (and sometimes better) ideas.

TURNING LEMONS INTO LEMONADE

After you have measured and analyzed performance and identified opportunities for improving performance, you can (or may be expected to) share this information with your donors, funding organizations, boards, agency leaders, or other stakeholders. Unfortunately, the prospect of doing this is what keeps some managers from being entirely candid about performance (and even measuring performance in the first place). Very often, they fear that their stakeholders will not understand or appreciate the significance of the information about performance. Although this fear may be well justified (more than a few public health programs have been defunded or restructured prematurely because decision makers failed to understand their performance), it can rob managers of a terrific opportunity to improve performance.

One way to alleviate fears about the consequences of sharing information about performance is to pair it with the story behind it. Doing so will greatly increase the likelihood that your audience will understand why performance isn't as good as expected or desired. It also may spur your audience to do their part to improve that performance. Furthermore, if you also present the opportunities or options you have identified for improving performance, you will point your audience to ways in which they can help.[28] Each of these outcomes will do a lot of good for a program, service, or process that isn't performing as well as expected or desired, so embrace the opportunity to share information about performance, and do it candidly, even when that performance is disappointing.

- Figure 1 in this guide can be used as a template to show performance data in conjunction with the story behind the performance.[29]

If taking these steps won't ease your fears about presenting information about performance, try one or more of the following additional strategies. Although they require more time and effort to implement than sharing the story behind performance and the options you have identified for improving that performance (information you already have if you completed earlier steps in the RBA process), these additional strategies can make a very powerful case for a program or service even in the face of disappointing performance.

- **Calculate Social Return on Investment (SROI),** which is the financial value of a program's economic and social impacts. For a relatively straightforward introduction to SROI, see https://www.thebalance.com/using-sroi-to-show-your-nonprofit-s-impact-2501977.[30]

28 Presenting the opportunities or options you have identified for improving performance may also result in praise for having a systematic performance improvement process in place!

29 For an example, see https://embed.resultsscorecard.com/PerfMeasure/Embed/164969.

30 Readers who have responsibility for community development programs may also want to explore the Low Income Investment Fund's Social Impact Calculator (http://www.liifund.org/calculator/), which enables users to calculate the social/financial value of selected community development projects including affordable housing, early childcare and education, K–12 education, and community health centers.

- **Present information about a program's performance in conjunction with the information about the roles that other programs and organizations play in achieving the program's objectives.** This can clarify for your audience what a program should and should not be accountable for.
- **Present information about a program's performance in conjunction with performance for the community, county, or state in which the program is operating.** This may enable you to show that your program's performance is as good as it can be given where it operates. It may also enable you to show your program's contribution to the performance for a larger area of interest.
- **Share a story that evokes powerful images and emotion.** If you think some of the members of your audience may not find the story behind performance sufficiently persuasive when you share it through data, try adding imagery that will make an emotional connection with them. In addition to bringing your audience around to your point of view, a story that includes vivid imagery in words or pictures is more likely to linger in their minds than a story supported by data, even when the data is compelling. Because members of your audience remember your story, they are more likely to repeat it and thus become promoters and champions of it – roles they may never have anticipated taking on before.[31]
- **Invite your audience to participate in the process of choosing the best ideas for improving performance.** Although sharing the opportunities/options you have identified for improving performance enables you to move your audience in the direction you would like them to go, you might consider going a step further by inviting the audience to refine or add to your ideas. If you do so, they will be less likely to think that you really just want them to "rubber stamp" your plans. You also may end up with better ideas!

PRESENTING PERFORMANCE DATA THAT IS UNDERSTANDABLE, CONVINCING, & MEMORABLE

Of course, no information about the performance of a program or service stands a chance of spurring its stakeholders to take action on its behalf if they do not understand this information or find it convincing. Unfortunately, **managers often don't (or can't) find the time to learn and apply best practices for presenting information, especially information in the form of data. Here is a selection of internet sites that cover these practices well:**

- NHS Education for Scotland has created a **general guide to visualizing (mostly numerical) performance data.** This guide explains factors that aid visual perception. It also provides an overview of some principles of design, and it reviews the basic types of data displays. It can be found at https://learn.nes.nhs.scot/3694/quality-improvement-zone/improvement-journey-measurement/presenting-data.

31 Use this strategy cautiously and sparingly. Some individuals may be put off by attempts to "tug at their heartstrings," so use these devices primarily when you know your audience is receptive to them.

- On her web site, data visualization expert Ann K Emery provides **detailed instructions for displaying data in different ways**, including these:
  - Presenting data "as-is" vs. presenting it with interpretation (http://annkemery.com/four-storytelling-strategies/)
  - Building a set of slides using storyboarding (http://annkemery.com/storyboarding/)
  - Developing a one-page annual report (http://annkemery.com/developing-annual-report/)
  - Using the best visualization tool of all time — the brain (http://annkemery.com/best-visualization-tool/)
  - Creating specialized charts, such as unit charts, circle charts, and span charts, without paying for the services of a design firm (https://depictdatastudio.com/tag/dataviz-challenge/ and http://annkemery.com/book/)
- In addition, Emery and fellow data visualization expert Stephanie Evergreen have created **a data visualization checklist.** It can be found at http://stephanieevergreen.com/updated-data-visualization-checklist/.
- BetterEvaluation.org has created **a summary of tasks, options, and approaches for reporting (and advocating for the use of) evaluation findings** that includes advice for choosing the right medium and tips on design. It can be found at http://www.betterevaluation.org/en/plan/reportandsupportuse. For a briefer review of approaches for reporting findings, see http://strengtheningnonprofits.org/resources/e-learning/online/analyzingdata/default.aspx?chp=3.

PRESENTING PERFORMANCE INFORMATION FOR OTHER PROGRAM MANAGEMENT ACTIVITIES

By now you know that RBA is much more than a process for measuring program performance. Among other things, RBA also

- fosters regular program monitoring,
- enables its users to identify the underlying and sometimes hidden causes of program performance problems,
- links program measurement directly to the identification of options for improving program performance,
- fosters better collaboration between a program and its partners, and
- can spur a program's key stakeholders to take action on the program's behalf.

For these (and other) reasons, **many people who use RBA think of it as a highly effective approach for carrying out a wide range of program management activities and not just for measuring and improving program performance (as important as that activity is!)**.

Here are some resources that can help you use RBA for some of these other program management activities:

- **Guidance for making the best use of performance measures in a grant application** can be found at https://clearimpact.com/3-17-how-do-we-use-performance-measures-in-writing-grant-applications/.
- **Guidance for using RBA in a meeting of agencies or other organizations that have responsibility for different parts of a service system** can be found at https://clearimpact.com/3-18-how-do-we-use-performance-measures-to-improve-cross-agency-service-systems/.
- **Guidance for using RBA as a framework for results-based planning** can be found at https://clearimpact.com/2-18-how-do-we-present-a-results-based-plan-to-the-public-to-political-leadership/.[32]
- Guidance **for preparing a progress report organized around an RBA process** can be found at https://clearimpact.com/2-21-how-do-we-report-on-progress/. [33]

KEY POINTS
STEP 6: SHARING AND PRESENTING INFORMATION ABOUT PERFORMANCE

Fears about the consequences of sharing information about performance are sometimes justified. One way to alleviate these fears is to pair performance information with the story behind it.

Offering opportunities or options for improving performance can give an audience some direction, and it may even persuade them to do their part to help improve performance.

Sharing additional types of information — such as the roles that other organizations play in accomplishing a program's objectives and performance data for the larger community, county, or state in which a program or service operates — will encourage stakeholders to have appropriate expectations for it.

Some excellent resources for learning how to present complex information that is understandable, compelling, and memorable are readily available online.

RBA is a highly effective approach for carrying out a wide range of program management activities including planning, partnership development, and grant and report writing.

32 For an example, see https://nmhealth.org/publication/view/report/2176/ (from New Mexico's Department of Health).

33 For two examples, see http://humanservices.vermont.gov/ahs-results-scorecard/ahs-results-scorecard (from Vermont's Agency of Human Services) and see http://raguide.org/progress-report-sample-results-based-accountability/ (from Mark Friedman and the Fiscal Policy Studies Institute).

SELECTING THE BEST IDEAS FOR IMPROVING PERFORMANCE

Most public health and human service program managers will not have the resources to implement every idea that might improve performance. How can they determine which one(s) are best? **Although no method for making such decisions is perfect, the better ones are designed to bring up the factors that will determine whether implementation of the idea will be successful.** The better methods also carve out time and space for engaging staff, clients, and partners in the selection process. Besides improving the chances that these key stakeholders will commit to implementing the ideas they have had a hand in selecting, engaging them in the process may also bring to light important, missing details about ideas that they are asked to consider.

Probably the most widely used method for selecting the best idea(s) from a larger group of options is criteria-based decision making. In the more structured versions of this technique, each of the options under consideration is scored on a number of criteria. If the criteria differ in importance, a numerical weight can be assigned to each of them so that their scores can be prorated before adding them up.[34]

- ASQ has created detailed guidance for implementing one type of criteria-based decision making that employs a matrix to list and weight criteria and evaluate options against them. For more information, see http://asq.org/learn-about-quality/decision-making-tools/overview/decision-matrix.html.
- The Public Health Foundation has developed an **Electronic Prioritization Matrix** that will automatically generate weights for any number of criteria based on a series of paired comparisons between them. For more information, see http://www.phf.org/resourcestools/Pages/Electronic_Prioritization_Matrix.aspx.

Three additional techniques for selecting the best idea(s) are worth noting:

- **The PICK (Possible, Implement, Challenge, and Kill) Chart** is another technique for scoring/rating options, but it only uses feasibility and impact as the criteria for scoring. A key feature of this technique is that it visualizes the scores. Post-It® Notes representing the options are placed on a four-quadrant matrix with potential impact on one axis and feasibility on the other. For more information, see http://www.phf.org/resourcestools/Documents/PICK_Chart.pdf.

34 Feasibility and potential impact are often selected as the most important criteria, but other criteria sometimes rise to the top instead. One of these is consistency with the values of the organization/staff/partners. Another is cost (although cost is sometimes integrated into the feasibility criterion). Originality may also be an important criterion if numerous efforts to fix a performance problem have not succeeded in the past.

- **Multi-Voting** (called "dot voting" when participants use colored dots as their votes) is a technique for identifying the most popular options in real time. Typically, each voter has the freedom to assign all of their votes to one option or split them among several options Two appealing features of this technique are the discussion and quiet thought that usually precede voting. The discussion time provides an opportunity to ask clarifying questions, while the quiet time provides an opportunity to process information. This technique is often used to narrow a large set of options down to a smaller set for further discussion. For more information, see http://www.phf.org/resourcestools/Documents/Multivoting.pdf.

- **Paired Comparison Analysis** is useful when a relatively substantial number of options are being considered, and few or no objective criteria for ranking them are available. It simplifies the task of ranking options by comparing each option to each other option to determine which one is preferred or more important. This comparison is repeated for every possible pair of options (usually in a table), with the preferred option receiving a score. The scores are totaled to create a list of options in order of importance. For more information, see https://www.mindtools.com/pages/article/newTED_02.htm.

Other techniques can be used to evaluate specific options (especially the more ambitious ones) in more depth:

- **Six Thinking Hats** allows an idea to be considered from a variety of viewpoints: positive, negative, emotional, rational, creative, and process oriented. In addition to enabling a thorough consideration of an idea, it is also useful for anticipating how other people will react to an idea. For more information, see https://www.mindtools.com/pages/article/newTED_07.htm.

- **Starbursting** quickly brings up the many important questions that need to be answered for an idea to be implemented successfully. For more information, see https://www.mindtools.com/pages/article/newCT_91.htm and https://business.tutsplus.com/tutorials/starbursting-how-to-use-brainstorming-questions-to-evaluate-ideas--cms-26952).

IMPLEMENTING THE BEST IDEAS FOR IMPROVING PERFORMANCE

Implementing ideas that require little or no additional funds or involve only a minor change to a process, service, or program is often easy, especially when the process, service, or program is fully within the control of just a small number of people. But implementing costly ideas, or ideas that involve a substantial change to a program/service/process that is highly visible or part of a larger program, can be tougher, even when the objective is to

improve performance. Such ideas may die on the vine unless they have substantial support, especially the support of top management, funders, and partners in agencies or other organizations that play a role in the program's success.[35]

Gaining the support of top management is critical for the success of many change efforts, especially those that require time and money. Many organizational development experts have written a great deal about ways to get support for an idea from top management, and some of these ways could be used to gain support for ideas identified through an RBA process. Getting support for the RBA process itself, on the other hand, might require a different approach. Here is a link to some strategies for getting top management's support for RBA (though many of them can be used to sway top management toward almost any idea or proposal): http://raguide.org/building-senior-leadership-support-for-results-based-accountability/.

Funding is a common obstacle to implementing ideas that require it. The list at this link suggests some alternative funding sources that aren't always considered: http://raguide.org/2-14-how-do-i-finance-a-results-based-plan/.

Even though **partnerships** are a tool in nearly every program manager's toolbox, they still can be difficult to establish and sustain because they require time, a high level of trust, a respect for other points of view, and sometimes a willingness to relinquish full control. There is no magic formula for establishing or sustaining a successful partnership, and in some cases, gaining the support of a program's partners may be more difficult than gaining the support of top management or a program's funders, despite wide agreement among the partners about the program's importance. Consequently, it may be useful to focus attention on the partners that will have the greatest impact on performance and the biggest role to play in improving it.

KEY POINTS
STEP 7: MOVING FROM TALK TO ACTION

The most effective methods for selecting ideas that can improve performance are designed to bring up the factors that will determine whether implementation of the idea will be successful. The best methods also engage major stakeholders (staff, clients, and partners) in the selection process.

Success in implementing an idea that requires a lot of resources, or an idea that involves a substantial change to a highly visible or large program, is much more likely if the idea has support from three stakeholder groups: top leadership in the organization operating the program or service, funders of the program or service, and partners in agencies or other organizations that play a role in the success of the program or service.

35 See Zachary, Brutschy, West, Keenan, & Stevens (2010) for an in-depth case study of the factors that were important for one municipality (Santa Cruz County, CA) to move from talk to action as part of an RBA process.

FINAL COMMENTS AND SOME WORDS OF ENCOURAGEMENT

In 2014, former CDC Director Thomas Frieden wrote an essay entitled *Six Components Necessary for Effective Public Health Program Implementation* (Frieden, 2014). In this essay, Frieden argued that successful public health programs share a number of characteristics. They (a) deliver a focused set of interventions supported by evidence; (b) monitor and analyze performance and use the findings to improve performance in real time; (c) establish effective partnerships; (d) maintain political commitment; and (e) communicate accurate and timely information to the health care community, decision makers, and the public to bring about behavior change and engage civil society.

In other words, successful public health programs define their role and purpose, measure their performance against this purpose, and use the data they collect on performance to figure out how to improve. They often include their stakeholders and partners in these activities, but they are also quick to address resistance, cynicism, fear, and misunderstanding. Successful public health programs always remember that their ultimate objective is to improve the health and well-being of the people they serve.

This guide has given you methods, tools, and tips for accomplishing all of these things — not just for programs but for services and processes as well. If you apply these methods, tools, and tips to a program, service, or process in your organization, its odds of being successful and effective are excellent. And you will thrive too. By adopting the practices in this guide, you'll measure what matters, identify ideas that will really work, establish more effective partnerships, and strengthen your ability to influence and inform. In addition, you will foster a culture of continuous learning in your organization – something that could have an impact that goes far beyond the impact you were seeking when you opened this guide for the first time.

REFERENCES

Bradt, D. A. (2009). Evidence-based decision-making (part II): Applications in disaster relief operations. *Prehospital and Disaster Medicine, 24*(6), 479-492.

Coalition for Evidence-Based Policy. (2012). *Rigorous program evaluations on a budget: How low-cost randomized controlled trials are possible in many areas of social policy*. (). Retrieved from http://coalition4evidence.org/wp-content/uploads/uploads-dupes-safety/Rigorous-Program-Evaluations-on-a-Budget-March-2012.pdf

Derose, S. F., Asch, S. M., Fielding, J. E., & Schuster, M. A. (2003). Developing quality indicators for local health departments: Experience in los angeles county. *American Journal of Preventive Medicine, 25(*4), 347-357. doi:S0749379703002083 [pii]

Dilley, J. A., Bekemeier, B., & Harris, J. R. (2012). Quality improvement interventions in public health systems: A systematic review. *American Journal of Preventive Medicine, 42*(5 Suppl 1), S58-71. doi:10.1016/j.amepre.2012.01.022 [doi]

Frieden, T. R. (2014). Six components necessary for effective public health program implementation. *American Journal of Public Health, 104*(1), 17-22. doi:10.2105/AJPH.2013.301608 [doi]

Friedman, M. (2005). *Trying hard is not good enough: How to produce measurable improvements for customers and communities* Trafford Publishing.

Herzog, K., Cooper, J., & and Holmes, K. (2017). Results-based accountability (RBA): An innovative framework to empower collaborative evaluation and continuous improvement. Retrieved from https://aea365.org/blog/tre-tig-week-results-based-accountability-rba-an-innovative-framework-to-empower-collaborative-evaluation-and-continuous-improvement-by-keith-herzog-jennifer-cooper-and-kristi-holmes/

Lee, A. (2013). *Results-based accountability (RBA) & connecticut state government.* (). Washington DC: Center for the Study of Social Policy.

Livingood, W. C., Sabbagh, R., Spitzfaden, S., Hicks, A., Wells, L., Puigdomenech, S., . . . Wood, D. L. (2013). A quality improvement evaluation case study: Impact on public health outcomes and agency culture. *American Journal of Preventive Medicine, 44*(5), 445-452. doi:10.1016/j.amepre.2013.01.011 [doi]

Macintyre, L. (2016). *Stop spinner your wheels: Using RBA to steer your agency to success* Fourth Quadrant Publishing.

Ng, E., & de Colombani, P. (2015). Framework for selecting best practices in public health: A systematic literature review. *Journal of Public Health Research, 4*(3), 577. doi:10.4081/jphr.2015.577 [doi]

Opinion Research Services. (2011). *Evaluation of results based accountability.* (Evaluation Report). Cardiff, Wales: Chronic Conditions Management Demonstrator Programme.

Perry, C. L., Komro, K. A., Veblen-Mortenson, S., Bosma, L. M., Farbakhsh, K., Munson, K. A., . . . Lytle, L. A. (2003). A randomized controlled trial of the middle and junior high school D.A.R.E. and D.A.R.E. plus programs. *Archives of Pediatrics and Adolescent Medicine, 157*(2), 178-184.

Petrosino, A., Turpin-Petrosino, C., & Finckenauer, J. O. (2000). Well-meaning programs can have harmful effects! lessons from experiments of programs such as scared straight. *Crime and Delinquency, 46*(3), 354-379. doi:10.1177/0011128700046003006

Schilder, D. (1997). *Overview of results-based accountability.* (Brief). Retrieved from https://1r65612jvqxn8fcup46pve6b-wpengine.netdna-ssl.com/wp-content/uploads/2013/03/evaluation-exchange-family-research-report.pdf

U.S. Department of Education. (2004). *The impacts of regular upward bound: Results from the third follow-up data collection.* (). Washington, D.C.: Retrieved from https://www2.ed.gov/rschstat/eval/highered/upward/upward-3rd-report.pdf

Yokum, D., Ravishankar, A., & Coppock, A. (2017). *Evaluating the effects of police body-worn cameras: A randomized controlled trial.* (). Washington, D.C.: Government of the District of Columbia. Retrieved from http://bwc.thelab.dc.gov/TheLabDC_MPD_BWC_Working_Paper_10.20.17.pdf

Zachary, D., Brutschy, S., West, S., Keenan, T., & Stevens, A. (2010). Connecting data to action: How the santa cruz county community assessment project contributes to better outcomes for youth. *Applied Research in Quality of Life, 5*(4), 287-308.

Zakocs, R., Hill, J. A., Brown, P., Wheaton, J., & Freire, K. E. (2015). The data-to-action framework: A rapid program improvement process. *Health Education & Behavior : The Official Publication of the Society for Public Health Education, 42*(4), 471-479. doi:10.1177/1090198115595010 [doi]

ADVISORY GROUP

Gayne Barlow-Kemper, MS
Early Head Start Director
Kennedy Krieger Institute
100 N. Chester Street
Baltimore, MD 21231

Thomas Higdon, JD
Volunteer Program Manager
Maryland Food Bank
2200 Halethorpe Farms Road
Baltimore, MD 21227

Jackie Range, MPH
Executive Director
Back on My Feet Baltimore
1017 E. Baltimore St. B
Baltimore, MD 21202

Gail Nutter Reid, MSW, LCSW-C
Director of Community Outreach and Engagement
The Family Crisis Center of Baltimore County
10 Dunmanway
Dundalk, MD 21222

Tulha Siddiqi
Program Director
Charm City Clinic
2222 Jefferson Street
Baltimore, MD 21205

APPENDIX B

COMMONLY USED AND ACCEPTED PERFORMANCE MEASURES FOR HEALTH AND HUMAN SERVICE PROGRAMS

NAME/TITLE & (SOURCE)	DESCRIPTION	TOPIC AREAS COVERED
CHRONIC DISEASE INDICATORS (CDC)	A set of 124 indicators (201 measures) in 18 topic groups that enables public health professionals and policymakers to retrieve uniformly defined state and selected metropolitan-level data for chronic diseases and risk factors that have a substantial impact on public health.	• Alcohol • Arthritis • Asthma • Cancer • Cardiovascular Disease • Chronic Kidney Disease • Chronic Obstructive Pulmonary Disease • Diabetes • Disability • Immunization • Mental Health • Nutrition, Physical Activity, and Weight Status • Older Adults • Oral Health • Overarching Conditions • Reproductive Health • School Health • Tobacco
COMMUNITY HEALTH INDICATORS TOOLKIT (FIRST NATION, CANADA)	Measures, data sources, and data collection strategies for a broad set of categories. Measures are based on input from community members and local health providers.	• Healthy Lifestyles • Services and Infrastructure • Others
EXAMPLE PERFORMANCE MEASURES (Clear Impact)	Measures for each of the three major types of program performance (*How much did we do, How well did we do it, and Is anyone better off?*).	• Child Welfare • Juvenile Justice • Mental Health • Others
FROM VISION TO ACTION: A FRAMEWORK AND MEASURES TO MOBILIZE A CULTURE OF HEALTH ((RWJF)	A selected set of 41 national, evidence-based measures representing all aspects of well-being.	• Creating Healthier, More Equitable Communities • Fostering Cross-Sector Collaboration • Improved Population Health, Well-being, and Equity • Making Health a Shared Value • Strengthening Integration of Health Services and Systems

COMMONLY USED AND ACCEPTED PERFORMANCE MEASURES FOR HEALTH AND HUMAN SERVICE PROGRAMS

NAME/TITLE & (SOURCE)	DESCRIPTION	TOPIC AREAS COVERED
CATALOG OF PERFORMANCE MEASURES AND INDICATORS (GOVEX)	Over 2,000 performance measures in use by cities across the U.S. Can be sorted by city and focus (topic) area.	• Education • Health and Human Services • Housing • Parks and Recreation • Public Safety • Others
METRICS FOR HEALTHY COMMUNITIES (WILDER RESEARCH AND THE FEDERAL RESERVE BANK OF MINNEAPOLIS)	Measures, data sources, and logic models for community development and health organizations wishing to work together to address social determinants of health and improve community health and well-being.	• Affordable Housing • Childcare Center • Community Health Center • Fresh Produce Access • Physical Activity • Supportive Housing
NATIONAL QUALITY MEASURES CLEARINGHOUSE (AHRQ)	Evidence-based health care quality measures divided into Health Care Delivery Measures (intended primarily for clinical providers and insurance plans) and Population Health Measures (intended for the assessment of public health programs, community influences on health, or population health characteristics that may not be directly attributable to the care delivery system).	• Health Care Delivery • Population Health *Note: Both types of measures are further subdivided into measures of Quality, Efficiency (includes cost as well as quality), and a third group (entitled Related health care) that are not supported by evidence of quality of care.* *They can be sorted in other ways as well, including by setting (e.g., Community Health Care)*
OUTCOME INDICATORS PROJECT (URBAN INSTITUTE AND THE CENTER FOR WHAT WORKS)	Outcomes, performance measures, and data collection strategies for 14 program areas	• Emergency Shelter • Health Risk Reduction • Youth Mentoring • Others
QUALITY POSITIONING SYSTEM (NATIONAL QUALITY FORUM)	A search tool for finding NQF-endorsed measures by condition, care setting, population, level of analysis (community, county, etc.), care setting, disease/condition, and type (efficiency, quality, outcome, etc.)	• Affordable Care • Disparities • Effective Communication and Care Coordination • Health and Well-Being • Health IT • Palliative Care and End-of-Life Care • Patient Safety • Person- and Family-Centered Care • Prevention and Treatment of Leading Causes of Mortality • Others

COMMONLY USED AND ACCEPTED PERFORMANCE MEASURES FOR HEALTH AND HUMAN SERVICE PROGRAMS

NAME/TITLE & (SOURCE)	DESCRIPTION	TOPIC AREAS COVERED
VITAL SIGNS: CORE METRICS FOR HEALTH AND HEALTH CARE PROGRESS (IOM)	A streamlined set of 15 standardized measures (each with its own set of sub-measures), with recommendations for their application at every level and across sectors.	• Addictive Behavior • Care Access • Evidence-Based Care • Healthy Communities • Obesity • Preventive Services • Patient Safety • Unintended Pregnancy • Well-Being • Others

APPENDIX C

HIGH QUALITY, ALREADY COLLECTED HEALTH AND HUMAN SERVICE DATA SETS

NAME/TITLE & (SOURCE)	DESCRIPTION	TOPIC AREAS COVERED
AMERICAN FACTFINDER (US CENSUS BUREAU)	Data and data summaries from nearly 100 surveys and censuses, including the American Community Survey, the American Housing Survey, and the Population Estimates Program.	• Economic and Geographic Information • Housing • Population Demographics (age, race, income, commute time to work, home value, veteran status, and other important data • Others
CDC WONDER (CDC)	A single point of access to a wide variety of reports and numeric and statistical public health data.	• Chronic Conditions • Communicable Diseases • Environmental Health • Health Practice • Health Prevention • Injury Prevention • Occupational Health • Others
COMMUNITY COMMONS	Easier access to data from federal and state data warehouses, with various reporting tools and capabilities for creating GIS-generated maps.	• Crime • Demographics • Children and Families • Clinical Care • Food • Health Behavior • Health Facilities/Professionals • Health Insurance • Health Rankings • Health Outcomes • Household and Housing • Poverty • Veterans

HIGH QUALITY, ALREADY COLLECTED HEALTH AND HUMAN SERVICE DATA SETS

NAME/TITLE & (SOURCE)	DESCRIPTION	TOPIC AREAS COVERED
DATA2020 (HHS OFFICE OF DISEASE PREVENTION AND HEALTH PROMOTION)	Data and technical information related to the Healthy People 2020 objectives, divided into 42 topic areas with state-level data maps.	• Access to Health Services • Adolescent Health • Environmental Health • Injury and Violence Prevention • Mental Health and Mental Disorders • Nutrition and Weight Status • Physical Activity
DATABANK INDICATORS (CHILD TRENDS)	National trends and research on more than 100 key measures of child and youth well-being, including health, social and emotional development, income, education, family, and community demographics and characteristics. Also includes info about interventions that have been proven to accelerate progress on each measure.	• Child Maltreatment/Child Welfare • Early Childhood • Education • Families and Parenting • Health • Poverty and Inequality • Social and Emotional Development • Teen Pregnancy and Reproductive Health • Youth Development
HEALTH INFORMATION GATEWAY (HHS OFFICE ON WOMEN'S HEALTH)	State- and county-level data. This system also enables data break down into several demographic categories, including gender, race, and ethnicity	• Access to Care • Demographics • Healthy People 2020 • Infectious and Chronic Diseases • Maternal Health • Mental Health • Mortality • Prevention • Reproductive Health • Violence and Abuse

HIGH QUALITY, ALREADY COLLECTED HEALTH AND HUMAN SERVICE DATA SETS

NAME/TITLE & (SOURCE)	DESCRIPTION	TOPIC AREAS COVERED
SELECTED METROPOLITAN/ MICROPOLITAN AREA RISK TRENDS OF BRFSS (CDC)	Analyzed data from CDC's Behavioral Risk Factor Surveillance System (BRFSS) that can be used to identify emerging health problems for metropolitan and micropolitan areas with more than 500 respondents. Data for all areas from 2011 to present can be found here.	• Alcohol Consumption • Asthma • Cholesterol Awareness • Colorectal Cancer Screening • Diabetes • Disability • Exercise and Physical Activity • Health Status • Health Care Access • Hypertension Awareness • Immunization • Nutrition • Oral Health • Tobacco Use • Weight Classifications • Women's Health *Note: The categories available will vary from year to year based on questionnaire changes.*