Slide 42 – Section Four IDENTIFY DATA NEEDED TO ANSWER A LEARNING QUESTION

This section of the training course focuses on how to identify data needed to answer a Learning Question. By the end of this section, you will understand the following concepts:

- Data identification
- Categories of data
- Common methods to obtain data

Slide 43 - Identify Data Needed

Data identification is based on the type of questions asked, the purpose of the study, and leadership needs. The type of question asked influences the data needs.

For Process Related issues, data needs are addressed through Who, What, Where, When, Why, and How questions.

For Outcome Related issues data needs are addressed through questions concerned with Impacts, Effects, and Changes.

Key considerations for determining what data is needed include:

- Is there existing data or standard measures?
- Will new data or standard measures be required?
- What data sources are required?
- Is data currently available?

Let's discuss data and types of Questions a little further.

Slide 44 - Leverage Data to Answer a Learning Question

As we learned earlier in this course, A Learning Question provides a way of framing the challenge you are attempting to clarify based on evidence. It guides how you proceed with research and ultimately analysis. Answering the question is determined by the identification, access, and acquisition of data.

Answering the learning question *"Does the data exist to answer the Learning Question?"* helps to resolve whether the data needs to be accessed, acquired, or obtained.

If the data is within the line of business or VA, it can be accessed through internal protocols.

If the data exists outside of VA, it can be acquired via public sources, inter-agency agreements, or specific data requests.

Finally, if the data does not exist, obtaining it will require additional research and a new collection plan.

Remember to employ Multiple Sources in your Data Strategy.

After receiving Data, continue the process of refinement and analysis.

Next, we will discuss Categories of Data.

Slide 45 - Categories of Data

The Two Primary Categories of Data are Quantitative Data and Qualitative Data.
These categories may be a review for some participants, but it is important to reinforce the distinctions.

Quantitative Data includes:

- Numerical and statistically analyzed data.
- Connections between components
- It also addresses questions of who, what, and how many.

Examples of Quantitative Data include:

- The number of Veterans who received employment through transition programs
- The age of Veterans applying for education assistance

On the other hand, Qualitative Data is narrative or nonnumerical, provides insights into experience and behavior, and addresses questions of how and why.

Examples of Qualitative Data include:

- Veterans' beliefs about VA support organizations
- Reasons Veterans do not apply for housing assistance

Let's explore these concepts in a little more detail.

Slide 46 - Common Methods to Obtain Data

Methods to obtain data are related to the required category of data needed.

Common methods to obtain Quantitative Data include Polls, Tests, Surveys, Assessments, or similar standardized forms.

These methods are applied to selected groups and populations.

For example, Veteran populations like those from Operations Enduring Freedom or Operation Iraqi Freedom are groups that could be polled.

Statistical methods and techniques are used to analyze Quantitative Data. We will discuss these techniques again later in the course.

Common methods to obtain Qualitative Data include Interviews, Focus Groups, Observations, and Document reviews.

These methods involve the first-hand study. Interpretative techniques like content analysis, case study, and similar techniques are used to analyze Qualitative Data.

Note that sometimes both Quantitative and Qualitative methods may both be used to address complex Learning Questions.

Combining quantitative and qualitative data is called the "mixed methods" approach to analysis.

Now let's review what we learned in this section.

Slide 47 - Section Four Summary

In summary, data identification is based on the type of questions asked, the purpose of the study, and leadership needs.

- Process Related questions address who, what, when, why, and how many
- Outcome Related questions address impact, effects, and changes

Answering a Learning Question requires a data strategy that assesses acquiring, accessing, and obtaining data from internal or external sources.

Quantitative and Qualitative Data are two primary data categories.

- Quantitative data and methods are numerical and statistical

- Qualitative data and methods are narrative and interpretive

Now that we've reviewed the key points for this section, let's have a quick knowledge check to reinforce your understanding of these concepts.

Slide 48 - Knowledge Check Number Five

How is data identified to answer Learning Questions? (check all that apply)

A. Type of questions asked
B. Leadership needs
C. Randomly selected resources

Slide 49 Knowledge Check Number Five Answer

If you checked A, Types of Questions asked, and B, Leadership needs, you are Correct.

C is Incorrect - Data identification should be aligned to the types of questions, leadership needs, and study purpose. Resource selections are based on prioritization and not a random activity.

Slide 50 Knowledge Check Number Six

Methods for Qualitative Data include: (choose all that apply)

A. Focus Groups
B. Assessments
C. Document Reviews

Slide 51 Knowledge Check Number Six Answer

You are Correct if you chose A, Focus Groups, and C, Document Reviews.

B, Assessments is Incorrect, as Assessments and Tests are Quantitative methods.

Next, lets discuss the importance of the quality, validity, and relevance of data collected.

Slide 52 – Section Five EXPLAIN THE IMPORTANCE OF QUALITY, VALIDITY, AND RELEVANCE OF DATA COLLECTED

The final section of this training course focuses on the importance of quality, validity, and relevance of data collected. By the end of this section, you will understand how to build evidence from data.

Slide 53 - Building Evidence from Data

This section discusses some technical considerations when building evidence from data. There are five topics we will review.

The first is that evidence should be available when needed and your work is useful by tying the conclusions reached to when the evidence is needed for making decisions.

The second is to keep an open mind and try to avoid bias or letting preconceived notions influence the evidence.

The third is ensuring that you find relevant data and develop data samples large enough for meaningful inferences.

The fourth is building confidence by understanding what information already exists and testing your assumptions and conclusions.

And finally, developing and presenting a wide range of specialties and expertise to which you may need access.

Let's start with the first topic, timing.

Slide 54 - Building Evidence from Data

Remember that evidence should be available when needed.

Schedule your work based on key decision points, and inform those decisions using the best evidence you have available at the time.

Plan for improvements in quality and scope over time if needed: building knowledge is not a "one-and-done."

Also, evidence is not correct forever, things change. Plan to maintain or revisit both data and analyses over time. You can incorporate new data into the same analysis to confirm results or identify new trends. Over time, test the extent to which the forecasts or relationships continue to be accurate.

Slide 55 - Building Evidence from Data

Next is the fun part of working to improve evidence for decision-making at VBA.

Every question you examine will have a history. Decisionmakers and you may have a sense of the likely, reasonable answers to your Learning Questions. *Those intuitions and existing evidence are a valuable cross-check* and need to be understood.

Often, decision-makers have a sense of the world and relationships, but it's likely based on anecdotal information, such as their experiences in a previous job. The answers may also just *seem* reasonable. *Evidence might confirm the hunches were right, but now we have confidence that we are making evidence-based decisions. Recognize that if people say, "we knew that all along." They may not have "known" it to the degree we need for defensible decision-making.*

Sometimes, decision-makers believe a situation or relationship exists but do not know to what extent. Within what range is it true? Your analysis may provide a similar answer on average as the educated guess, but you can determine the degree of accuracy of the information. *Narrowing the range of uncertainty is a key function of improving the evidence supporting decisions.*

What if the results are a real surprise? What if they seem to contradict our hunches, or significantly disagree with existing information? In such cases, you must examine, as honestly as possible,

- Whether the new evidence provides better evidence than the existing information
  OR
- If there might be a mistake in the data, the analysis, or in the strength of the conclusions you drew

Either way, VBA has learned something valuable.

Slide 56 - Building Evidence from Data

Next, let's talk a little about getting the data you need for your analysis.

To convert data into evidence, find data where you can and develop samples large enough for meaningful interpretations.

Administrative data developed in VA is the easiest source of data to obtain and may be the fastest source, but they may not be as informative as needed.

Remember, you aren't alone — Administrative data from other agencies and even private sources, such as academia, may be available.

Match the data sought to the evidence needed. Decisionmakers may not need great precision. Maybe the decision-makers need to confirm whether the actual value of something is above or below a certain value that will trigger their decision. For example, they may need to know whether a certain problem confronts a small handful of Veterans or is a problem for many. If just a few are affected, perhaps additional information could be provided in a Frequently Asked Question Action page, but a program redesign may be warranted if over 5 percent or so of veterans are affected. Data collection and analysis are usually greatly simplified knowing more about the likely decision point.

Good sampling practices may limit the need for additional data to reasonable levels, perhaps a few hundred or a few thousand data points. Precision increases rapidly with sample size, and developing good evidence rarely requires a census of affected veterans or even a large sample of thousands. However, there are restrictions on asking for information from more than nine entities and nine data points can rarely answer an interesting question. The next two slides cover when you must get approvals for new data collection projects.

Slide 57 - Paperwork Reduction Act

As you develop your approach to improving evidence, you may determine that available administrative data is insufficient to support the scope or quality of evidence you need. You may need to develop new, original data.

Generating new data may require getting approvals. In particular, the Paperwork Reduction Act may apply.

The Paperwork Reduction Act was enacted to reduce the administrative burden on individuals, businesses, and institutions from the collection of information by or for the federal government.

New information that can only be obtained from the public that is critical to evaluating or studying your program can still be obtained, but you need to plan ahead. Approvals take time and require compelling justifications and publication in the Federal Register for public comment. VA's Paperwork Reduction Act officer can help guide you through the process.

Most agencies estimate six to nine months for PRA clearance which includes at least three months of public comment. You can start the comment period in the Federal Register and consider public comments while simultaneously developing the more detailed request for OMB review.

Slide 58 - Paperwork Reduction Act

Certain actions trigger getting approvals through the Paperwork Reduction Act process.

For example, seeking the same information or asking the same question of 10 or more entities in a formal survey or research study usually requires approval from the Office of Management and Budget. This applies even if responding is a request, not a requirement. It applies to most questions posed to the public but also applies even to federal employees and contractors if the question is not directly related to their work-related duties.

There are a few exceptions which do not trigger PRA such as collecting information from federal employees or military personnel as a part of their job, general requests for public input directly tied to their interactions with the VA, such as an open-ended "Tell Us About Your Experience" question, and discussions at public meetings.

Other laws may also be relevant if seeking information from other jurisdictions, such as from state personnel, or if you are seeking certain kinds of personal information. VA's Office of General Counsel can help guide you through these issues and others related to collecting new data.

Actions that do not trigger the Paperwork Reduction Act include collecting information from federal employees or military personnel as a part of their job.

Slide 59 - Building Evidence from Data

The previous slides discussed obtaining data. Of course, having data isn't enough--we have to convert it into information and evidence. We have already briefly discussed data quality, but here are a few related concepts that make data useful.

One useful concept is that creating information and evidence from data requires grouping or comparing data that are similar enough to be meaningful. Basically, it makes better sense to compare a basket of one kind of apple to a basket of similar apples, rather than to a basket of different apple varieties or a mixed-fruit basket.

There are common challenges when ensuring data can be combined or compared. Try to answer questions like these:

- Is the data close enough in time?
- Is the data from the same or similar enough population?, and
- Is the data close enough in quality?

The overall question is: Can you defend to decision-makers combining and comparing the data in the way you intend?

One example is how you might use data from the time when the pandemic affected most of our operations and the lives of Veterans trying to engage with VBA programs. Data from this period may be less useful to combine with earlier years for projecting trends, or as a comparison for judging the efficiency of operations.

Slide 60 - Building Evidence from Data (Continued)

A second concept is that providing information about the distributions of values usually provides better insights than looking at an average—often it is the "tails" of distributions that matter most. For example, understanding more about the range of the slowest or least successful 10 percent of interactions may provide more actionable evidence than the average. An average may change very little and mask a bigger change in the extreme values.

A third concept is about planning for the work. There will be some tough work needed in the background. In particular, allow plenty of time for "cleaning" the data to make it comparable and useful. Analysis requires that data is all consistently formatted for analysis, is mostly complete, is accurate with no obvious outliers, and is well documented so that others can verify or replicate your work. You will need to watch out for extreme values that can impact your summary and statistical measures.

Also, try to prove the extent to which the data is well suited to your purpose. Any use of the data for evidence implies the data serve as useful proxies for the future. These two questions often help.

- Is the future likely to be much different than when the data was collected?, and
- Do you expect the future to have similar cause and effect relationships as the data you collected?

Be careful as your worldview will impact how you see the data, correlations of cause and effect, and how it fits together. It will also help you to explain why the data and the resulting evidence are important. That sense of how the world works—that is, *what* causes *what*— is your mental model that explains how the evidence helps in decision-making.

Now, let's move on to a discussion of how to build evidence from data.

Slide 61 - Building Evidence from Data

How can you demonstrate that the evidence you provide is worth decision-makers' confidence?

There are two ways to increase confidence in your work. Confidence in the evidence increases dramatically when you can demonstrate you have taken a "big picture" view of the evidence.

The first is to demonstrate support for the same findings from multiple lines of evidence. That is, show that data and evidence from different sources produce similar results, or at least do not strongly contradict your findings. Show that data and evidence from the same sources over time produce similar results. And show that data and evidence using different methods produce similar results. Confirming conclusions using multiple lines of evidence is very powerful.

The second way to increase your confidence in your work is by demonstrating that hard questions have been asked. Try to ask ***and answer*** the hard questions that the users of the evidence may ask. For example, try to disprove your hypothesis and explore other factors or reasons that explain the results.

Slide 62 - Building Evidence from Data

The field of data collection, analysis, and presentation is large and has many specialties. One person is unlikely to perform all the required functions well. One beginning assumption is that you may need to obtain the skills you need to get started, grow at least some of the skills you need internally in your office over time, and maintain the skills you need over time. While this course does not provide background on all the specialties that may be needed or useful, here are a few that may come into play.

The most obvious expertise that may be needed is quantitative statistical methods, possibly including correlation, regression analysis, analysis of variance, and Monte Carlo simulations.

Depending on your data used, another expertise that may be needed is qualitative research methods, possibly including surveys, content analysis, narrative analysis, and grounded theory.

Regardless of the approach taken, another expertise that you should consider is data visualization to help explain and tell the story of the evidence.

There are specialized tools to which experts in these fields may need access and you may need to arrange for support. These tools may include Excel, Tableau, Power BI, Nvivo [pronounce: En-Viv-O], SAS, SPSS, R, S-Plus, or others.

Slide 63 - Example of Descriptive Data

Simple descriptive statistics can help leadership better understand performance.

Here, we have an example of legacy appeals, illustrating a week-over-week reduction. The information is useful but lacks context about the bigger picture.

By adding additional aspects, we can identify a broader story for the decision-maker.

First, we add the target; identifying the target helps the leader understand the end goal in mind.

But this still lacks context. Are we progressing fast enough to meet the goal?

By adding an expected progress line, using just Excel, we can demonstrate how many appeals should have been cleared in order to stay on track.

Let's move on to a summary of this section of the course.

Slide 64 - Section Five Summary

Evidence is not accurate forever, things change. Plan to maintain or revisit both data and analyses over time.

Match the data sought to the evidence needed.

Generating new data may require getting approvals. In particular, the Paperwork Reduction Act may apply.

Summarizing information from data requires grouping Like-with-Like to identify patterns and generate other statistics.

Confidence in the evidence increases dramatically when you can demonstrate you have taken a "big picture" view of the evidence.

Now that we've reviewed the key points for this section, let's have a quick knowledge check to reinforce your understanding of these concepts.

Slide 65 - Knowledge Check Number Seven

You realize the findings of your study differ from what some of the program managers expected.

Which steps could you take to help prepare for the discussion with leadership?

A. Let the evidence speak for itself
B. Recheck the methodology and calculations
C. Present factors may have changed which would explain the difference between expectations and the results of the new study
D. Include a comparison of the quality of the study with existing sources of information in terms of currency, sample size, methodology, and relevance—examine the body of evidence as a whole
E. Leak the results to the public so the managers cannot disavow the results
F. Present the Bottom Line Last—Detail the thoroughness of the methodology before offering the conclusion

Slide 66 - Knowledge Check Seven Answer

There is not a single correct answer for this scenario. Choosing to recheck the methodology and calculations of your analysis, option B, or presenting the bottom line of your analysis last, option F, are both valid responses. Options C and D are also both appropriate responses to this question. However, options A and E are not correct responses. The purpose of this section of the course was to help you develop your evidence-building capabilities and to think like an analyst. When interpretations of data differ, an analyst has multiple options to pursue. They can reexamine their dataset or make clear in their findings that the evidence is incomplete.

Slide 67 - Apply Concepts to a Hypothetical Situation

We have covered quite a few concepts. Let's try to apply them to a couple of Learning Questions based on a hypothetical situation.

Slide 68 - Apply Concepts to a Hypothetical Situation

The hypothetical situation is based on a new benefits program to encourage Entrepreneurial Veterans. The program intends to partially correct a market failure where Veterans have less success getting start-up business loans because the link between the skills, they acquired during their service is less understood by loan officers than regular private sector experience. The new program is a loan program funded by a self-supporting revolving fund.

The two primary goals of this new program are to expand Veteran employment and business ownership.

VA is to begin accepting loan applications within 1 year and begin issuing loans within 18 months.

Further, the program will be re-evaluated by Congress in 5 years based on the experience of the program and alternatives VBA has considered.

Slide 69 – Learning Question: How Effective Is Business Creation?

In our hypothetical program, Congress will review the program in 5 years. To allow a program redesign and recommendation to Congress in 5 years, VBA requires a program evaluation completed well before that, with enough time to develop recommendations to Congress on any program changes.

In this hypothetical situation the Learning Question could be stated as: How effective has the program been in creating new Veteran-owned businesses?

More specifically, within 4 years, we would want to answer this question—

- To what extent has the program created new Veteran-owned businesses that would not otherwise have been created without this program?

Let's run through a set of considerations to evaluate the new program.

Slide 70 - Key Considerations: How Effective is Business Creation?

What category of evidence is most needed in this hypothetical situation?

- Policy Analyses
- Program Evaluation
- Performance Measurement
- Foundation Fact Finding

Slide 71 - Key Considerations: How Effective is Business Creation?

As posed, this Learning Question is about evaluating the effectiveness of the program against its stated goal. It will require comparing what did happen against our best estimate of what would have happened if the program had not been in place.

If the question were phrased as comparing two or more alternative ways of achieving the same outcomes, it would have been a policy analysis question. Performance management would have focused on established metrics, and foundational fact-finding would have focused on understanding the underlying relationships of factors affecting the success of the program.

What are some considerations of such a program evaluation?

The effectiveness of the program is not just counting the new businesses--some of these entrepreneurial Veterans may have started businesses anyway. Maybe they would have focused more on smaller, self-employment-type businesses that required less capital than was available through the new program.

Estimating the difference between what did happen and what would have happened is challenging. Sometimes this problem is comparing what has happened to a baseline—meaning what happened previously under similar conditions.

Sometimes it is estimated by creating and tracking a "control group" for comparison. For example, maybe Congress allowed a pilot program in some states, and Veterans in the unaffected states could be tracked for comparison. A study to see how the Veterans who applied for, but did not receive the loans, could be used as a "control group."

Another approach to estimating what would have happened without the program would be to model the likely behaviors and create a simulation of what would happen under the same conditions.

There are also qualitative ways to estimate what might have happened. For example: you can interview or observe participants in your study. Interviewing and re-interviewing over time Veterans who didn't apply for the loans, and those who applied but did not receive them could provide insights into what might have happened without the program.

Slide 72 - Key Considerations: How Effective is Business Creation?

Next, let's think about the timelines.

What are the timelines? Is the example a

Short-term objective or Long-term Objective

Slide 73 - Key Considerations Learning Question: How Effective is Business Creation?

In our hypothetical example, this is clearly a longer-term objective with a study needed in 4 years. But to perform this evaluation well will require early planning and probably continuing data collection to have the information necessary to perform the study.

Next, let's consider how the data is likely to be obtained.

Slide 74- Key Considerations Learning Question: How Effective is Business Creation?

Obviously how the data is obtained depends a great deal on how the study is designed.

How is the data likely to be obtained? (More than one option can be selected)

- Administrative data within VA
- Administrative data within government
- Data from academia
- New Data

Slide 75 - Key Considerations Learning Question: How Effective is Business Creation?

If you selected **administrative data within VA**, **administrative data within government** or **new data**, you are correct. There will be critical data generated by running this hypothetical program. Obtaining data from other agencies on business creation will be important for this evaluation as well. It is also likely that some new data will be needed, such as from interviews or surveys. Also, studies done by outside groups (such as academia) could be directly useful.

Slide 76 - Key Considerations Learning Question: How Effective is Business Creation?

Next, what are the most important skills that may be needed in this study.

- Project Management
- Interview and collaboration meeting skills
- Analytical skills
- Knowledge of VBA operations and program
- Legal and Regulatory

Slide 77 – Key Considerations Learning Question Number Two: How Effective is Business Creation?

The best answers here are project management, interview and collaborative meeting skills, and analytical skills. Answering a program evaluation-type question certainly requires a skilled project manager to plan and orchestrate a multi-year effort. The nature of this program indicates that

surveys and interviews will likely be a key source of information, so those skills would likely be required. Developing high-quality evidence requires appropriate analytical skills. This may include the production and analysis of qualitative data derived from interviews and other sources.

The other two areas may come into play, but the previous skills would be the focus in answering this type of learning question.

Slide 78 - Key Considerations Learning Question: How Effective is Business Creation?

After participating in this exercise, you should have a better idea of how to use the concepts covered during this course to answer learning questions.

For additional information and guidance on Evidence-Based Policymaking, you may access the Evidence Act Toolkits designed by the GSA Office of Evaluation Services.

Additional resources from Think-Tanks and Other Non-Governmental Organizations are provided.

Slide 79 – Summary

Let's review some key concepts discussed in this course.

Selecting the correct category of evidence to answer a Learning Question is the key to developing evidence. Multiple categories of evidence will often be required to answer a Learning Question.

The first consideration for designing a study is identifying the evidence and resources needed to answer the Learning Question.

Answering a Learning Question requires a data strategy to acquire or access data from internal or external sources.

Evidence is not accurate forever and both data and analyses evolve over time.

Generating new data may require approvals and the Paperwork Reduction Act may apply.

Finally, building and presenting evidence may require specialized skills and resource.

With this information, you now have a better understanding of how to develop evidence for decision making.

Thank you for participating in Advanced Evidence Based Policymaking.

If you are interested in additional evidence based training materials and resources.

These additional training resources are available for your use.