Quality Improvement Basics: Data Analysis

Slide 1 Objectives
After completing this module, participants will be able to:

- Explain the key basic steps in data analysis.
- Describe the purpose of descriptive and diagnostic data analysis.
- Discuss what data stratification is and why it can be helpful.
- Define mean, median, mode, and range in analyzing data.

Slide 2 Data Analysis
Once you have collected your data, it is time to start reviewing and analyzing it. You've probably heard the expression 'slicing and dicing' data…that's what data analysis is all about.

You'll want to look at the data from different angles and through different lenses, and this is typically done with tools like Microsoft Excel, Apple 'Numbers,' Google Sheets, or other similar business and consumer applications. If you are more advanced or have a dedicated team member with data analysis and reporting expertise, they may use tools such as Tableau, Power BI, RStudio, SAS, or others. Some electronic health record vendors are also offering modules or add-ons to conduct data analytics. It's a good idea to check in with others, such as your superusers, IT department, or EHR vendor, to see what capabilities are available in your organization.

Slide 3 Data Analysis Process Steps
Data analysis aims to provide information that will help you make informed decisions.

While the quality improvement process is continuous, the data analysis process involves five linear steps that can be repeated as many times as necessary to drive improvement projects.

Step 1: Define data goals and objectives.

- First, we must define what we are trying to understand or achieve with the data. For example, if you want to understand what is happening with a process or an outcome. This requires input from all people involved in the QI project and may need to be revisited and revised throughout the analysis process. Often, data goals are too broad or ambitious and must be narrowed down or broken into multiple sections.

Step 2: Collect the data.

- When defining your data goals, make sure you consider what types of data are accessible to you and how it will be collected. In the QI Basics Data Collection and Monitoring module, we addressed who is responsible for collecting the data, how and when it will be collected and where it will be stored.

Step 3: Clean the data.

- Raw data is often messy and inconsistent. Thoroughly cleaning the data ensures it is easier to work with in the next two steps of the process. Clean data is also more accurate and reliable. Cleaning the data means to:

- Check for consistent units of similar variables (for example, if you are looking at weight measurements, check to see if they are recorded consistently in the metric OR in the imperial system, and decide if you will use decimal points or round up to whole numbers)

During this step you will also:

- Remove any irrelevant or duplicate data,
- Look for big gaps or missing data
- Reconcile any inconsistencies and
- Format appropriately.

Step 4: Analyze the data.

- This means working with your data to start to glean useful information. Using the data analysis tools mentioned earlier, review, manipulate, break down, and test the data to better understand what the data is portraying. Look for trends, patterns, outliers, or variations that might tell a story.

Step 5: Interpret and share insights.

- Evaluate the data from your analysis and determine what relationships, trends, and interactions are taking place. Make recommendations and solicit feedback from team members about ways that the data could be impacted differently to better achieve your data goals.

Let's dive deeper into step 4: Analyzing the data.

Slide 4 Methods of Data Analysis

There are many ways to conduct analysis and better understand your data. The main types of analysis used in health care quality improvement are descriptive, diagnostic, inferential, and predictive.

- Descriptive data analysis summarizes the data to highlight anomalies, trends, and underlying issues. It helps to answer the question: "What happened?"
- Diagnostic data analysis identifies causes and helps to answer the question: "Why did this happen?"
- Inferential and predictive analysis utilizes statistical models and techniques to determine causations and predictions in the data. This answers the question: "What is likely to happen in the future?"

For this module, we will be focusing on descriptive and diagnostic analysis. It's important to note, since we are not conducting scientific research, we cannot conclude specific causations. However, we can use analytic tools to identify correlations between variables that help us pinpoint useful insights.

Slide 5 Describing Your Data

When describing data, the two main objectives are summarizing the data in useful ways and identifying trends and anomalies within the dataset.

When we talk about quantitative data, there are some terms that help us describe the values we've collected. As we'll learn in the QI Basics Data Display Methods module, it's important to communicate and tell the story about the data that describes what is occurring in the process or what is happening with the measure we are focused on.

Here are a few summary statistics terms that are useful to know about numeric data sets:

This project is supported by the Health Resources and Services Administration (HRSA) of the U.S. Department of Health and Human Services (HHS) as part of an award totaling $740,000 with 0% financed with non-governmental sources. The contents are those of the author(s) and do not necessarily represent the official view of, nor an endorsement, by HRSA, HHS or the U.S. Government. (June 2023)

- Mean – average of all numbers. This is the sum total of all collected data points divided by the number of data points. Let's take five example numbers: 3, 5, 8, 10, and 50. The total is 76, Divide that by the number of data points, five, and you get the mean: 15.2
- Median – the middle value (50% of data is above, 50% is below the median), organize your data sequentially…lowest to highest, and then determine which value lies in the middle. In this case eight is our Median since three and five lie below it and 10 and 50 are above it.
- Mode – the value that occurs most frequently. Our data set of 3, 5, 8, 10, and 50 has no mode as each number occurs once. If our data set were 3, 3, 5, 8, 10, and 50, then three would be our mode as it occurs most frequently in that data set.
- Range – the difference between your highest and lowest value. Our range is the highest value minus the lowest value or 50-3 which is 47.

Slide 6 Diagnostic Analysis

Diagnostic analysis involves discovering anomalies (good or bad) that are not fully understood or explained in the current state. Diagnostics then drills into the data to look for patterns and determine potential relationships that could be causing those anomalies.

You can identify trends or anomalies by making comparisons across groups, categories, or benchmarks, and visualizing how the data compares to other groups or how categories of data compare to each other. You will also be able to display and identify any outliers or potential inaccuracies within the dataset.

Evaluating data changes over time is helpful when studying how different process changes impacted the data at different stages of our project. Time analysis is also useful to make educated guesses about what could happen in the future.

Stratification, or arranging or classifying data into smaller groups, or "strata" allows you to identify interactions and relationships within the data.

You can examine possible relationships between factors and find patterns and correlations that point to potential causes. This may allow you to identify variables that are essential in driving the performance of your data goals (these are called key performance indicators or KPIs). If you already have established KPIs, you can examine their relationship to different factors and their interaction with desired outcomes.

Slide 7 Stratification

Part of data analysis is stratifying your data or looking at your data through different lenses or strata.

Examples of stratifying health care data include looking at data according to:

- Time of day
- Day of week
- Site of care
- Care providers
- Procedures
- Patient characteristics (such as age, gender, race, ethnicity, and more)

Sometimes your data may already be stratified into the needed level of detail based on how it was collected. The attribute of time may have been collected by year, month, week and then displayed in the raw format it was collected in, such as the day it was collected. If you collected at a more granular

This project is supported by the Health Resources and Services Administration (HRSA) of the U.S. Department of Health and Human Services (HHS) as part of an award totaling $740,000 with 0% financed with non-governmental sources. The contents are those of the author(s) and do not necessarily represent the official view of, nor an endorsement, by HRSA, HHS or the U.S. Government. (June 2023)

level, you may have reason to display the time of day or hour that it was collected. By stratifying from a time perspective, you may discover variations and differences based on, for example, the shift that performed a certain task.

Stratifying data enables you to look at the data through different lenses…almost like a camera zooming in and out of a landscape. If you are viewing your data set from a very high, summarized level, you may need to 'zoom' in and stratify your data at a more granular level to recognize patterns or differences. For example, you can look for variation by time of day, clinician, or patient characteristics such as race and ethnicity.

Slide 8 When to Stratify

In general, you will stratify data if you suspect that some characteristic may provide needed insights into whatever you are measuring. In addition, perhaps you suspect that, for example, the types of admissions vary by patient zip code and certain geographical regions may have more complex diseases. Or, in another example, perhaps staff adherence to a standard protocol of care differs during weekends or late at night. Stratification can help you better understand your data and can aid in designing the most effective and impactful intervention to improve quality.

When data is aggregated, it can mask important differences. For example, an aggregate readmission rate may look great, but when you break it down into other strata (by diagnosis group, for example) you may find big differences that better highlight areas for improvement. You may think you're doing quite well when there are actually some large gaps or disparities. As you collect and study your data and are initially seeking to identify areas for quality improvement, stratification can reveal opportunities to focus on.

Before you begin your actual data collection, consider any categories or characteristics you'll need to stratify and ensure you are collecting that level of detail.

Slide 9 In Summary

- Key steps in the data analysis process include defining what we are trying to understand or achieve with the data, collecting, cleaning, and analyzing the data to provide useful information to guide decision making.
- Descriptive data analysis summarizes the data to highlight anomalies, trends, and underlying issues. It helps to answer the question: "What happened?" Calculating the mean, median, mode, and range of a data set can be helpful. The mean is the average of all numbers, median is the middle value in the sequence, mode is the number that occurs most frequently and the range is the difference between the highest and lowest value.
- Diagnostic data analysis identifies causes and helps to answer the question: "Why did this happen?" This can be accomplished by looking for trends or anomalies, stratifying the data, or identifying relationships in the data.
- Stratification is sorting into distinct groups so that patterns can be seen. If all data are lumped together, it can be difficult to determine the data's meaning.

This project is supported by the Health Resources and Services Administration (HRSA) of the U.S. Department of Health and Human Services (HHS) as part of an award totaling $740,000 with 0% financed with non-governmental sources. The contents are those of the author(s) and do not necessarily represent the official view of, nor an endorsement, by HRSA, HHS or the U.S. Government. (June 2023)