Quality Improvement Basics: Data Analysis and Data Display Methods Transcript

Slide 1:
Welcome to the Quality Improvement Basics course 'Data Analysis and Data Display Methods' module.

Slide 2:
Our topics in this module will cover:

- How to conduct Data Analysis
- Displaying your data: Charts, Graphs & Tables

Please note that there are several references to the PDSA, Plan Do Study Act improvement tool which precedes this module and the assumption is that you have a basic understanding of that.

Also, before we start, please open the related documents for this module (tools, templates and any samples) which are available on the web page where you found this module link. It will help you to have those ready for quick reference as screenshots of the documents may not legible on your screen.

Slide 3:
Data Analysis

Slide 4:
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 it from different angles and through different lenses and this is typically done with Microsoft Excel, Apple 'Numbers', Google Sheets or similar business and consumer applications. If you are more advanced or have a dedicated team member who understands data analysis and reporting, there may be using tools such as Tableau, R, or others. These are much more expensive and require special training. Also, EHR vendors are starting to offer modules or add-ons that can accomplish data analytics. Be sure to ask your super-users, IS or IT department or vendor if that might be an option as this is easily overlooked.

Slide 5:
Part of data analysis is the exercise of looking at your data through different 'lenses' which will lead to different conclusions about your data.

Stratification determines the level of detail or breaking the data down into characteristics of the data set.

Examples of stratifying healthcare data include looking at data according to:

- Time of day
- Day of week
- Site of care
- Care providers
- Procedures
- Patient characteristics

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

Stratifying the data is enabling yourself and your audience to look at the data through difference lenses…almost like a camera lens 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, such as the example of seeing variation according to which shift performed a certain task.

Slide 6:
In general, you will stratify the data if you suspect that some characteristic of the data may provide needed insights into whatever you are measuring. In addition, perhaps you suspect that, for example, the types of admissions vary by zip code– that more complex diseases may be more likely to produce errors. 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 that you'll need to stratify and make sure you are collecting that level of detail.

In summary, you stratify your data …
- To more fully understand the data, you have
- To help design an intervention to achieve your quality improvement goals

Slide 7:
When we talk about the data we've collected, there are some terms that help us describe the values we've collected (quantitative types of data) that go beyond simply stating an average. As we'll learn in the data visualization section coming up, it's important to be able to communicate and tell the story

about the data as it describes what is occurring in the process or with the measure that we have chosen to focus on.
Here are a few terms that are useful to know about numeric data sets:

- 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 5 numbers: 3, 5, 8, 10 and 50. The total is 76, Divide that by the number of data points, 5 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 8 is our Median…3 and 5 lie below it and 10 and 50 are above it.
- Mode – the value that occurs most frequently. Out 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, 3 would be our mode as it occurs most frequently
- Range – the difference between your highest and lowest value (Max-Min=Range) our range is the highest value minus the lowest value or 50-3 which is 47.

Slide 8:
Displaying your data: Graphs, Charts and Tables

Slide 9:
How you want to communicate your collected observations will be your guide in selecting the appropriate tool, whether that ends up as a table or some type of visual graph or chart. There is an entire field and science around data visualization, this term used to describe how we present and tell our stories with data. Experts in this field frequently use the term 'story telling' to convey the idea that the purpose of data visualization is to provide background, perhaps a series of events that lead us to a conclusion or action…all part of the data story and the tools (our graphs charts and tables) and a means to that end.

As you think about telling your story with data…

- Display data with the end in mind (what goal are you trying to achieve by creating graphs, charts and tables…you are trying to communicate and tell a story and most likely persuade or convince someone to take actions or draw some sort of conclusion)
- Decide whether to summarize or display all the data you have collected (charts graphs and tables enable us to condense a large story into a much shorter, summarized one)
- Form of data (number, percent, precision, etc.)
- Who is your audience? (have you tailored your story to meet their needs at their needed level of detail?)
- And be sure to choose a Method that best tells your story or is appropriate for your type of data

As we learn about the options in the upcoming slides, you'll start to form some ideas about what types of graphs, charts and tables will work best for you particular improvement story and how to match the visualization tool to the type of data you have and what you want to 'say' with it.

Slide 10:
Graphs and Charts are very similar, and the terms are often used interchangeably, however, a graph is a depiction of a mathematical function or formula whereas a chart is a more simple term to cover the graphical depiction of data.

Graphs and charts:

- Show comparisons between variables
- They Show patterns, trends or relationships
- And they Convey complex data relationships in a quick manner

Tables are used to display individual values and enable the viewer to see the relationships between those data points. Tables show quantitative (numeric) information in a row and column format with cells that are populated with the actual data we have collected. If the data story you are trying to tell must relay specific values, rather than larger trends, then a table would be the correct tool for your need. Depending on your audience, you can start with a graph or chart and then, after that, include a table which provides specific values if you need to tell your story with numbers at a more detailed level.

Slide 11:
Each type of graph, chart or table was designed with a certain mission…it will do a great job at telling your data story if you use it for it's intended purpose.

Pie charts are best used to show how a whole is broken into its parts when there is a very limited number of parts.

Scatterplots do a great job conveying relationships between two variables

Bar charts, line graphs and tables are intended for displaying comparison between values, such as categories or groups of data

…and Histograms and Scatterplots are the right tool for presenting distributions of data

Regardless of which tool you use, be sure to keep it simple and don't overpopulate your graph, chart or table with data. Your goal is to condense the data into a brief data story that has some sort of direct point or call to action for your audience. If you overpopulate your page with data, lines, legends, color etc.…your audience will focus on interpretation which will detract for the point you're trying to make. Less is more when it comes to data visualization.

Slide 12:
Here are several examples of charts and graphs which display the same data using different methods. The data shown here depicts the percent of patients broken down into ethnic representation for the entire patient population for a specific facility.

The top two bar charts simply have the x and y axis reversed, the lower right line graph substitutes a line for a bar and the lower left pie chart displays the categories according to what percent they represent based on the total for all observations.

As you learn about the different methods of displaying the data, and how each graph, chart and table is used, you will draw some conclusions about which tools may be best suited to tell your 'data story' and support your Quality Improvements case.

Slide 13:
One of the simplest approaches to sharing your data is to compile and share it in a table format with rows and columns, rather than presenting it in a graphical format which will cover next. However, your audience will immediately be drawn into the exercise of trying to interpret the data you are showing rather than understanding the bigger story you are trying to tell. Therefore, showing the actual values you have collected should only be done if discussing those values is needed. Consider a chart or graph as a preferred method to help your audience absorb the story faster.
If you are convinced that you need to present the actual values, you'll need to make decisions about how much or how little data to share. Don't, of course, overwhelm your audience with too much data with endless rows and columns. If you can define the 'story' that you want to tell, this will help you narrow that table to a subset of your data or perhaps an aggregated or summarized view.

Our example here shows the distribution of ethnic populations for a fictitious facility along with the percentages represented by each group, the percent diagnosed with hypertension and the actual number of diagnosed patients by population. As you look at this table, think about how long it is taking you to make sense of this data and we'll then display the same data in a bar chart. You'll start to appreciate the difference it makes for your audience when using the appropriate data visualization tool.

Also, those who use tables frequently in their presentations (for the right reasons) recommend that if you want your audience to compare specific values, do so in a column format as we've done here, rather than a row across the page.

Slide 14:
For quality improvement, histograms are used to present the spread, and the shape of the data distribution. That is, where is it centered, how spread out is it, and whether it is symmetric or skewed in one direction or the other.

Histograms are used to show distributions of variables and plot a range of quantitative data. Bars can be reordered in bar charts but not in histograms.

In our example here the ranges are broken into groups of 10-minute wait time categories…1-10, 11-20, 21-30 minutes and so on. This ranges are displayed on the horizontal or 'X' axis. The vertical or 'Y' axis shows the total counts for each particular range, in this case the count of patients reporting their wait times within a specific range of minutes.

By looking at the graph, we can see that the majority of patients waited somewhere between 41- and 70-minutes last year (the blue bars) and in the current year patient wait time were significantly reduced (in orange) from about 1-30 minutes.

Slide 15:
Bar charts are used to compare categories of observed data. In the example here, the categories represent the counts of patients by ethnic categories who have a hypertension diagnosis and also the count of those with a body mass index greater than 25.

The height of the bar represents, in this case, the count or amount of any given category. You can reorder the categories to emphasize a certain point or trend with the data set, greatest to least for each category, for example.

Bars can be reordered in bar charts but not in histograms (recall that our histogram categories were sequential 1-10, -11-20, etc.). Typically, the bars of bar charts have the same width as the focus is on the height or y axis.

What our bar chart tells us very quickly, is that the total patient count is very close (looking at the vertical Y axis) for the number of patients with a hypertension diagnosis and for the count of those with a body mass index greater than 25. While not conclusive, our data tells us that there may be some correlation between hypertension and a BMI over 25.

Slide 16:
Line graphs, or line charts, are often used to display a change over time to identify trends, patterns, and exceptions. Our line graph here shows a monthly comparison of newly diagnosed patients with hypertension (the blue line). and those that received a 2nd blood pressure check on their first visit (the orange line).

Line graphs can show patterns in data clearly by showing the values over time. In many cases, the slope or trend of the line can be used to help make predictions beyond the data we've collected out into the future. Line graphs can easily display one or more continuous datasets (the blue and the orange line each represent a data set here). However, displaying more than about five lines on the same graph can make the data difficult to interpret and makes the graph a bit cluttered.

Slide 17:
An alternative to bar charts is pie charts. The primary strength of a pie chart is that they indicate a part-to-whole relationship. However, you do want to be careful to limit the number of categories displayed. A maximum of 3 or 4 categories is an acceptable limit and be aware that the visual comparison is difficult if the contribution to the whole is very small (for example, 1-5 % out of 100% is almost imperceptible in a pie chart).

In the examples on the right, the upper pie chart has 8 categories which is too many and comparing the categories is difficult. In the lower right pie chart, we have reached our maximum of 4 categories and you can see that it is easier to make the relative comparisons.

However, it is better to use histograms as the viewer can more readily compare the height of the bars to see the differences between categories.

Comparing the size of different "pieces" is more difficult that comparing the size of rectangles that are part of histograms or bar charts. If you go back 5 slides to the slide entitled 'Same data displayed differently' you'll see that pronounced difference between using a histogram than a pie chart and how much easier it is to make the comparisons. In the field of data analysis and analytics, pie charts are a controversial topic and you'll find that some experts highly discourage the use of them.

Slide 18:
Scatter Plots require that you have two continuous variables, and when you plot each "pair" of data on the x (horizontal) and y (vertical) axes the graph displays the relationship between these data points. The example here shows the relationship between patient reported wait times and the number of MDs and NPs working at the facility on various days during the measurement period.

In general, it will tell you the association between two variables, and under certain circumstances may give you insight into a possible causal relationship, although cause-and-effect cannot be proven by scatter plots. You may identify a correlation (meaning there is consistency between the two variables…when one goes up the other follows in the same direction or vise versa)…. but this does not necessarily prove causation between the two.

When we use scatterplots, we:

- Test for possible cause and effect
- Keeping in mind that this not prove a cause & effect relationship exists
- It can verify whether a relationship exists, and we may want to use that relationship to further investigate the process we are intending to improve
- A cause and effect relationship will be verified only after the improvement is tested and results are studied

On a scatter plot, you plot the exact data based on where it should appear based on the scale of your X and Y axes. It is the concentration and frequency of the data that will lead to insights and conclusions from this type of diagram.

Slide 19:

This concludes our overview of charts, graphs and tables as well as our two modules that cover data in our QI Basics course. As you've learned from this section, using data is an expansive topic and there is much more to learn.

Remember to use an appropriate amount of data needed to tell your story as you consider who your audience will be and how you want to tell your story with data. Keep your story simple and impactful as using an overabundance of data and charts can easily distract or loose your audience entirely. In the final module of the course 'Pulling it all Together" , you'll see an example of how Using Data is integrated into a quality improvement project.

Slide 20:

Thank you for taking time to learn about Data Analysis and Data Display Methods as part of the QI Basics course. Please join me for the final module in the course: Pulling it all Together.

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 $625,000 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. (September 2019)