healthwatch

# Quantitative data analysis guidance

August 2021

# About this guide

As part of your work at Healthwatch, you routinely collect large amounts of data which you need to sort through to understand what it means and make clear evidence-based decisions. This guide will help you recognise the steps you need to follow when analysing quantifiable data to improve your data analysis skills and produce impactful reports. There are also top tips for analysing quantitative data and how best to represent your analysis visually.

# What is quantitative data analysis and why it is important?

When you can count, measure and express data as numbers, you carry out quantitative data analysis. Contrary to qualitative data analysis, it generates results that are concise, rigid and defined. It can help you answer questions such as "how much" or "how many", and you can visually represent your findings using graphs and charts. Typically, at Healthwatch, you would collect quantitative data via surveys; however, you could also collect such data via your feedback and signposting activities.

Quantitative data analysis is much quicker than qualitative data analysis and involves less complicated processes. It generates only a specific range of values, allowing you to generate unbiased information from the population more easily.

**Example of quantitative data analysis at Healthwatch:**

Between January – March 2021, you have 642 pieces of feedback from the public telling you about their experiences using local health and social care services. You notice that 378 of those relate to people's experiences of accessing NHS dental services.

When you analyse this data, you can see that 59% [(378 ÷ 642) x 100], or nearly three in five people who contacted you during this period, had something to say about NHS dentistry. This data gives you a clear indication that access to NHS dental care is a priority issue in your area. You can use this evidence-based information to plan your course of action, for example, raising the problem with your local NHS England colleagues.

Note that the figures of 642 (total number of feedback), 378 (number of feedback about dentistry) and 59% (proportion of feedback related to dentistry) are concise, rigid and defined – you cannot report them as any other value.

# Before you analyse quantitative data

Consider the following:

1. What is your question, or what do you want to find out? Remember that your question(s) should be measurable, clear and concise.
2. What value(s) you would measure that will help you answer your question?
3. How will you measure it? For example, what is your time frame or unit of measure?
4. Are there any factors that you need to consider, e.g. demographic characteristics of survey respondents?
5. How you will collect and organise the data?

**For example:**

You want to find out about COVID-19 vaccine uptake among the minority ethnic people in your local community.

You design a survey asking people whether they have had the vaccine. You also ask respondents about their ethnicity.

You can measure the proportion of respondents from minority ethnic groups who have answered "yes" to your question. This will be out of the total number of people who have responded to your survey.

You might want to look at the proportion of specific ethnicities or age groups within an ethnic group.

In this case, you must ensure that you collect data from as diverse groups of people as possible.

Once you have collected the data, you can organise it on an Excel spreadsheet with information from each ethnic group on a separate tab to make analysis and comparisons easier.

# Steps in quantitative data analysis

## 1. Data management

- **Create a database where you will store the data**. You can create it by using software such as Microsoft Access or even Microsoft Excel. The former is a dedicated software used to create databases; however, it is subscription-based. On the other hand, Excel is easily accessible for most and can create simple databases. You can use column headers in Excel to mark data fields and apply the filter function to extract relevant data from the database when needed.
- **Add data to the database and make sure that they map correctly**. When you add data, you must ensure it matches the correct field, for example, making sure that the age of respondents is under the age field. You must also ensure that the format of the data remains consistent throughout the database. For example, if you collect gender data as "male/female", do not use "M/F" interchangeably, as this can lead to incorrect analysis, especially in large datasets where it might be challenging to see and correct such discrepancies.
- **Check for incomplete or dissimilar data**. It is essential to ensure that you arrive at the correct conclusion following your analysis and analyse the data correctly. For example, if information is missing from specific fields, your analysis might present a partial picture that could be different from when you have all the fields complete. Similarly, if you have dissimilar data in a field (e.g., M/F instead of male/female), you will end up with incorrect counts and percentage calculations.
- **Clean the data by removing any errors or discrepancies**. For example, in Excel, you can apply a filter on a field and check for differences in the data. So, if you use a filter on gender data and spot "F" instead of "females", you can filter the data and convert all Fs into females.
- **Store your data in a secure location and follow data governance policies, as appropriate**. We have produced guidance about data governance you can refer to and learn more from.

### Top tips for a good data management strategy

- Only collect data that you don't have and plan to use.
- Make sure you collect good quality data. The information that you collect should be accurate, complete, reliable, relevant and up to date.

- Automate data cleaning where possible, e.g. by using VLOOKUP formulas in Excel.
- Have dedicated staff who are responsible for managing the data.
- Create your data policy outlining what you need to consider about collecting, storing, using and archiving data. You can find out more about this by reading the various guidance produced by us.
- Keep your data management strategy up to date with the latest technology and legislation. Find out more in our top tips about storing your data and keeping it secure.

## 2. Understand the variables in your data

Before you start analysing the data, it's essential to go through it and understand what the different variables are. Doing this will enable you to work out what you can and cannot measure.

For example, if you have collected data about service areas and the ease of access to help and support, you can tell which services people have struggled to access most. However, if you haven't collected the age of the respondents, you will not be able to find out if access was more difficult for certain age groups.

## 3. Carry out a descriptive analysis

Descriptive analysis or descriptive statistics help you to describe the basic features of a data set. This enables you to summarise the information in the data and present it in a simple and meaningful way. Descriptive analysis is the most common type of quantitative analysis used in Healthwatch research.

There are four types of descriptive analysis outlined and explained in this guidance. However, the most common ones used in Healthwatch research are the measures of data frequency and central tendency, as described below:

### Measure of frequency

Measuring frequency helps to establish how often a value occurs in a dataset. You can do this by using the following measurements:

- **Count** – this tells you the total number of observations in a category, e.g., the number of respondents to a survey.
- **Frequency** – this is the number of times a particular value appears in the data, e.g., the number of male respondents to a survey.
- **Percentage** – this tells you how a value within the data relates to a larger or the total value. You should always represent this as parts per hundred. For example, if 35 out of 78 survey respondents identify themselves as "males", that is 45% [(35 ÷ 78) x 100].

### Measure of central tendency

This tells you how numerical values are distributed in your dataset in relation to the central position within that data set. It does not provide information about an individual value but rather a summary of your dataset. You can measure central tendency in the following ways:

- **Mean** – this is the numerical average of a set of values. To calculate the average, you will have to add up all the numbers and divide it by how many numbers there are.
- **Median** – this is the midpoint of a set of numerical values. To find the median, you will need to place the numbers in order and find the middle.

- **Mode** – this is the most common value among a set of values. To find the mode, you will need to place the numbers in order and identify the value the occurs most frequently.

For example, you have interviewed nine people as part of a research project. Their respective age in years is as follows: 26, 43, 37, 26, 29, 54, 66, 34, and 30.

The average (mean) age of your research participants is: (26+43+37+26+29+54+66+34+30) ÷ 9 = 38.3 years

Their median age is: 26, 26, 29, 30, **34**, 37, 43, 54, 66 = 34 years

Note: If you have an even set of numbers, you will need to find the middle pair and divide their sum by two to get the median value. So, if you had ten research participants, and their age was 26, 26, 29, 30, **34, 37**, 40, 43, 54, 66, the median will be 35.5 years [(34+37) ÷ 2)

The mode value for this dataset is 26 - **26, 26**, 29, 30, 34, 37, 43, 54, 66

## Top tip

It is important to note that while mean (or average) is the most common method of reporting the central tendency in Healthwatch research, the median is helpful to report as well, especially when there are extreme values (or outliers) in your data. This is because the mean takes all the values into account, and therefore, is easily influenced by the outliers. The median represents the true mid-point of your data as there is an equal number of points above and below it.

For example, you are researching how people's income influences their access to high-speed broadband. You survey over two hundred people who provide you with their annual income figures. You notice that majority of the respondents earn between £25k-£30k annually. However, a few earn less than £18K and two people earn more than £80k. If you want to correlate income with access to high-speed broadband, it is better to report the median income of your research participants in this case. If you only report the average income, there is a risk of skewing the real picture about affordability and access, as the mean will include both very high- and low-income values.

## 4. Carry out comparative analysis where relevant

Comparative analysis helps you to compare two or more variables in your dataset. Doing this allows you to find out information about one or all the variables you are comparing. You can also compare similar variables over time. Typically, at Healthwatch, comparative analysis must be carried out to find differences in people's experiences from different demographic groups, such as age, gender and ethnicity.

For example, you undertake a survey to find out what people think about the costs of NHS dentistry and how it impacts their decision to seek dental care. You collect their demographic characteristics and analyse:

- Percentage of males vs percentage of females who struggle to afford the costs

- Percentage of respondents aged under 25 years who have seen a dentist recently compared with those over 65 years

- Number of Black females who cannot afford the costs compared with their White counterparts

- The average income of people who prefer to see a private dentist compared with those who prefer to see an NHS one

### Other types of descriptive analysis

You may sometimes need to carry out other types of descriptive analysis depending on the kind of data you collect. Healthwatch research does not commonly use them.

### Measure of variation

Measure of variation helps you to understand how the values are spread in your dataset. Low variability means that values in your dataset are clustered around a point, while high variability means that the values are spread out. You can measure the variation in your dataset in the following ways:

- **Range** – this is the difference between the highest and lowest value in a set of values. An extensive range means high variability, while a small range indicates that your data has low variability in distribution. A range is a valuable measure when there aren't any outliers in the data, i.e., extreme values.

  You might want to report the range of certain values in your dataset when they are not too varied. For example, when you carry out research with young adults between 18- 25 years of age, you can report the age range of seven years (25-18).

- **Standard deviation** – tells you how far, on average, each value in your dataset lies from the mean. This tells you if a specific data value is 'standard and expected' or 'unusual and unexpected'. The larger the standard deviation, the more variable your dataset is.

### Measure of position

Measures of position allows us to see where a data value falls relative to other values within the dataset. This tells us whether the value is average, unusually high, or low compared to the other values. The three most common ways you can measure position are:

- **Percentile** - this is the value that divides your rank-ordered dataset, i.e., when the values in your dataset have been ordered from smallest to largest, into 100 equal parts.
- **Quartile** – this is the value that divides your rank-ordered dataset into four equal parts or quarters.
- **Interquartile range** – this is the middle 50% of your dataset.

## Top tips for analysing quantitative data

- First, carry out a descriptive analysis of your data to understand its core characteristics.
- Then undertake comparative analysis to check how the different variables fare against each other.
- You do not need to carry out all the types of analyses – it should depend on your data and what you want to find out.
- Focus on using the method(s) which provides the most significant impact on your reporting.

# Top tips for data visualisation

- When deciding how best to represent your data visually, think about its purpose, the intended audience and what information you think would be the best to present.
- Do not create data visualisation for every single piece of analysis that you do – this will confuse your audience and reduce the impact of your report.
- Bear in mind that your aim is to present the information from your analysis in a way that makes it easily and quickly clear for the person who is reading your report.
- Choose the correct type of data visualisation for your analysis, e.g., use a bar chart for quick comparisons, pie charts to show the proportional distribution of the variables you are measuring or line plots to represent a trend over time.
- Where possible, arrange your data in an increasing or decreasing fashion, such as in bar charts.
- Keep it simple – avoid unnecessary use of gridlines, bold illegible text and 3-D graphs.
- Use colours effectively, such as contrasting colours to highlight differences between groups or use shades of the same colour to depict waning intensity.
- If you know that you will print the visualisation black and white ink, then use textures instead of colours to differentiate the data.
- Include a zero baseline where possible to avoid giving a false visual representation.
- Label your data visualisation and make sure it is legible and easy to read. Place the legend where it is easy to read. Always give a title to your graph.
- You can read more about how to report your quantitative findings in this guidance.

# Other considerations and further reading

- When designing your research methodology, consider if you need to collect quantitative data and how that will help you to answer your research question. You may not necessarily need to collect numerical data in all cases. In fact, on some occasions, it is better to collect only qualitative data to find meaningful, nuanced information, for example, when trying to find out about the impact of inadequate post-natal support on new mothers during the pandemic. Read more about how to plan a research project in our guidance.
- Use mixed methods of data collection where relevant – this means you can collect both quantitative and qualitative data simultaneously. For example, designing surveys with a few free-text questions or undertaking a survey and then identifying participants for a focus group discussion to explore the details about a topic.
- When reporting your quantitative analysis, consider how you will present your findings of the target population. Be careful not to generalise - especially if the data collected is not from a representative group from the target audience. You can read more about how to get the right sample for your research in this guidance.