INTRODUCTION

The aim of this ‘How To’ guide is to provide advice on how to analyse your data and how to present it. If you require any help with your data analysis please discuss with your divisional Clinical Audit Facilitator who will be happy to help.

1. HOW TO ANALYSE DATA

Audit data comes in three different forms, ‘tick-box’, numerical or freetext. Each requires different methods of analysis, but in each case the aim is to establish which standards are being met (% compliance) and which are not (% non-compliance). If a standard is not being met you need to identify why and how practice can be improved to ensure that the standard is met in the future. You may also consider if there were other, acceptable reasons for the standard not being met, i.e. an exception not considered during the planning stage.

A. TICK-BOX DATA

It is likely that the majority of the data that you have obtained from your data collection form will relate to yes/no options or tick-box options from a specified list of alternatives. In such cases, it is usual practice to add up the number of answers recorded for each option and express the total as a raw number and as a percentage.

**EXAMPLE 1:**
- Sample size: 50 patients
- Audit criteria: All patients should attend a pre-operative clinic
- Question: Did the patient attend a pre-operative clinic?
- Results: Yes = 32 and No = 18.

A good way of expressing this data is:

- All patients should attend a pre-operative clinic. \( n=50 \)
  - Yes = 32 (64%)
  - No = 18 (36%)

The ‘n=50’ indicates how many patients were in the audit sample and is used to calculate the percentages, i.e. 32/50 = 64%.

It is important to remember that yes/no options do not allow for ‘not applicable’ answers. Taking the example used above, it is possible that certain patients did not meet the standard because they had an emergency operation. In this instance the answer to the question ‘Did the patient attend a pre-operative clinic?’ would have been ‘not applicable’. To reflect this the data can be expressed more accurately as:

**CALCULATING COMPLIANCE WITH CLINICAL AUDIT STANDARD**

\[
\text{Number of patients who meet standard} \times \frac{100}{\text{Number of patients to whom standard applies} - \text{Number of patients who meet any listed exceptions}}
\]
EXAMPLE 2:
- Audit criteria: All patients should attend a pre-op clinic
- Exception: emergency operation
- Results: Yes = 32, No = 13 and N/A (emergency) = 5

\[
\frac{32}{45} \times 100 = 71\%
\]

- 32 patients attended a pre-op clinic
- 18 did not, but 5 of these were emergencies (exception criteria)
- Therefore 32/45 (71%) met the standard

B. NUMERICAL DATA

Some of the data items you collect are likely to be numerical values, e.g. age, length of stay in hospital, blood glucose level etc. Lists of numbers like this can be summarised using measures of central tendency and dispersion:
- Measures of central tendency look at the middle/common values in a list of data items: the mean, median and mode.
- Measures of dispersion look at how spread the data is: the range.

MEASURES OF CENTRAL TENDENCY
The mean is the average value, calculated as: \( \text{Sum of all the values} \div \text{Number of values} \)

The table below shows data about length of stay (LOS) on three wards:

<table>
<thead>
<tr>
<th>Length of Stay (days)</th>
<th>Number of patients discharged</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ward 1</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
</tr>
<tr>
<td>4</td>
<td>18</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
</tr>
<tr>
<td>6</td>
<td>18</td>
</tr>
<tr>
<td>7</td>
<td>12</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
</tr>
</tbody>
</table>

For Ward 1, the mean is:
\[
\frac{(1\times4) + (2\times8) + (3\times12) + (4\times18) + (5\times20) + (6\times18) + (7\times12) + (8\times8) + (9\times3)}{4 + 8 + 12 + 18 + 20 + 18 + 12 + 8 + 3} = \frac{511}{103} = 4.96
\]

The mean LOS on ward 1 is 5 days (rounded to nearest whole day).

If the same formula were used to calculate the means for wards 2 and 3, you will find that for each ward, the mean LOS is 5 days. However, the mean is not always the best measure of central tendency.

The LOS for all three wards is illustrated on the graph below. The mean suggests that the data is the same for all three wards, however the graph indicates that this is not the case. The problem with the mean is what it does not tell us.
The data collected for Ward 1 is almost perfectly symmetrical, with the graph illustrating that the data follows the shape of a ‘bell curve’. Data that conforms to this shape is known as ‘parametric’ data. In this instance the mean is an appropriate measure of central tendency.

The data for Wards 2 and 3 is non-parametric. Their graphs do not form a symmetrical curve. Describing their notable features, Ward 2 has a significant proportion of patients with a LOS of 3 days together with a number of patients staying 9 or 10 days. Ward 3 has a peak LOS of 6 days. It can be seen that using the mean alone with non-parametric data is not very informative. The median and mode can help to convey the missing information.

The mode is the most commonly occurring value. For Ward 2 this is 3 days and for Ward 3 it is 6 days. This should be obvious from both the raw data and the graph. If the highest occurrence is shared by more than one value you could either state them all as modal values, or none. For example, if for Ward 3 there were 10 patients discharged on both day 5 and day 6 you could either say there were 2 modal values of 5 and 6, or that there was no mode.

The median is the mid-point of all the values. For Ward 2, we have data on 67 patients. If we made a list of LOS, placed in order from the lowest to the highest, the mid-point would be the 34th value i.e. there are 33 values below and above this. The 34th value relates to a patient who was discharged after 5 days, so this is the median. For Ward 3, we have data on 42 patients, i.e. there is no single mid-point. In this case, take the average of the 21st and 22nd value (there are 20 values below and above these two values). The 21st value relates to a patient who was discharged after 5 days and the 22nd value relates to a patient who was discharged after 6 days, so the median is 5.5 days (5+6 divided by 2).

Unless you are well versed in statistics, we would advise that you use all three measures of central tendency, or show the information using a graph. In general, quote median rather than mean for non-parametric data.

Not all lists of numerical data should be analysed in this way. For example, if your standard is ‘The patient will be considered medically fit for surgery if temperature <38°C’ and you collect a list of temperature data, it would not be meaningful to present the mean, median and mode temperature. What you are interested in here is the percentage of cases that met the standard i.e. the percentage of surgical cases with temperature <38°C.

MEASURES OF DISPERSION
As well as stating the mean, median and mode, it is also good practice to provide some indication of how spread the data is. The range states the lowest and highest values. In our example:

Ward 1 has a range of 1-9 days
Ward 2 has a range of 1-10 days
Ward 3 has a range of 1-8 days
A more subtle way of expressing dispersion is to use quartile range. This involves listing your values from lowest to highest, as per calculating the median, and then dividing the values into four equal parts or sub-ranges. The range you are interested in lies between the second and third quarter (or ‘quartile’).

So, for example, some more LOS data:

Ward C: 1 2 3 5 | 5 6 6 7 | 7 7 7 9 | 9 11 20 38

In this case the range is 1-38 days, but the quartile range is 5-9 days

The quartile range is useful in taking out outlying data (data some distance away from the mean/median/mode), as in the case of Ward C above and Ward 2 in our first example. Ward 2 has the largest range but a comparable quartile range to Wards 1 and 3:

Ward 1 - range 1-9 days; quartile range 4-6 days
Ward 2 - range 1-10 days; quartile range 3-7 days
Ward 3 - range 1-8 days; quartile range 4-7 days

### ANALYSING DATA AGAINST STANDARDS

If your standard statement was ‘Patients should be discharged by the end of their 5th day following surgery’, using LOS data for ward 2, you find that 45 out of the 67 discharged patients had a LOS of 5 days or less.

You would write this as 45/67 (67%) patients met the standard.

### C. FREE-TEXT DATA

If you include an open question in your data collection form, you will obtain free-text data. In order to analyse this data you should group comments into themes or categories, i.e. as if you were creating tick-box options for the data collection form. You might also want to consider reproducing some comments verbatim in your report, if they are particularly pertinent.

### DRAWING CONCLUSIONS

The end stage of your analysis is concluding how well the standards were met and, if applicable, identifying reasons why the standard was not met in all cases. These reasons might be agreed to be acceptable, i.e. could be added to the exception criteria for the standard in future, or will suggest a focus for improvement. In theory, any case where the standard (criteria or exceptions) was not met in 100% of cases suggests a potential for improvement in care. In practice, where standard results were close to 100%, it might be agreed that any further improvement will be difficult to obtain and that other standards, with results further away from 100%, are the priority targets for action. This decision will depend on the topic area, in some ‘life or death’ type cases, it will be important to achieve 100%, in other areas a lower result might still be considered acceptable.

### 2. DISPLAYING DATA

#### CONTINUOUS DATA

‘A set of data is said to be continuous if the values/observations belonging to it may take on any value within a finite or infinite interval. You can count, order and measure continuous data. For example height, weight, temperature, the amount of sugar in an orange, the time required to run a mile’.

[http://www.stats.gla.ac.uk/steps/glossary/presenting_data.html](http://www.stats.gla.ac.uk/steps/glossary/presenting_data.html)
Continuous data fits on a numerical scale. Examples of continuous data include:
- Temperature: 34°, 35°, 36°, 37°, 38°, etc.
- Days post-op: 1, 2, 3, 4, 5, 6, 7, 8, etc.
- Age: 16, 17, 18, 19, 20, 21, etc.

Generally when displaying continuous data, it would be appropriate to use charts that allow you to lay out this numerical scale and plot data against it e.g. scatter graphs, box and whisker plots, etc.

CATEGORICAL DATA

Categorical data is data that can be sorted according to non-overlapping (mutually exclusive) categories, whereby each subject in a sample can only fit into one category. For example:
- Gender: Male/ Female
- Age group: 16-20, 21-25, 26-30, etc.
- Standard met: Yes / No

The usual type of data resulting from a clinical audit is categorical data and would most commonly be represented as a bar chart or a pie chart.

PIE CHARTS AND BAR GRAPHS

Bar or pie charts are most commonly used in clinical audit to illustrate compliance with audit standards, however there are certain circumstances where one or the other is better.

- **Pie charts** - Show proportion, e.g. percentage compliance with standard. A pie chart would be inappropriate if, for example, it was being used to illustrate ‘type of treatment’ in cases where patients are likely to have had more than one treatment. In this instance the pieces of the pie would add up to the total number of ‘treatments’ rather than the total number of patients, which could be misleading and is not very meaningful in itself.
- **Bar/column charts** - Generally used to show frequency, e.g. number of patients seen by different staff: nurse, Specialist Registrar, Consultant, etc. For example if the audit criteria stated that all patients seen in A&E meeting certain criteria should be seen by a consultant, you might want to show what grade of staff saw the patients if it was not the consultant.
- **Versions of Bar charts (stacked or comparative)** - Show more than one standard/ question per chart, e.g. multiple results on one graph, bars divided to show percentages, etc.

**EXAMPLE 3: Pie Charts, Bar Graphs & Comparative Bar Graphs**

![Pie Chart Example](image1)

![Bar Graph Example](image2)
CREATING GOOD CHARTS

Good charts should focus on getting your message across, rather than creating fancy, and distracting, images. Clutter should be avoided and the charts clearly labelled.

REMEMBER: Do not use graphs just for the sake of it.

EXAMPLE CHART 1: 37/40 members of staff (93%) took personal protective equipment on domiciliary visits

A chart might be considered unnecessary to illustrate the above data. It is always important to consider whether or not people need to see a graphical representation of the data, in some cases simply expressing the data as ‘x/n (y%)’ is sufficient. In a project with a lot of standards producing a chart for every single one may confuse rather than clarify the results. People may not remember which image related to which standard.
EXAMPLE CHART 2: What’s wrong here?

What’s wrong?

- The 3D graph often leads to misleading graphs: because the bars are not up against back wall of chart, you could mis-read 25 and 15 as 24 and 14 or lower.
- The title needs more detail.
- The axes should be labelled. In this case the x axis might not need labelling according to what other titling is put on chart, but the y axis needs to be labelled to be meaningful.
- In this instance the legend is not needed, as there is only one data series.
- There is a lot of white space on the page. It looks unprofessional.

Revised version

- It has been changed to 2D chart.
- Titles have been added.
- The format of gridlines and background has been amended to add clarity.
- The sample size number has been added for quick reference.
- There is better use of the space available.
- The scale has been extended a little way past highest bar.

SUMMARY

- Consider whether or not you need to represent your data graphically to get your message across.
- Consider which type of chart will deliver your message in the clearest way.
- Consider what information you should include in order to answer your audience’s questions. For example, sample size, percentages if displaying real values (or vice versa). Importantly, avoid clutter.

CONTACT DETAILS/ USEFUL INFORMATION

CLINICAL AUDIT

- The UHBristol Clinical Audit website is available [online] via: http://www.uhbristol.nhs.uk/healthcare-professionals/clinical-audit.html
- Contact details for the UHBristol Clinical Audit Team are available from the Clinical Audit Central Office or [online] via: http://www.uhbristol.nhs.uk/healthcare-professionals/clinical-audit/contacts.html
- The full range of UHBristol ‘How To’ guides are available [online] via: http://www.uhbristol.nhs.uk/healthcare-professionals/clinical-audit/how-to-guides.html
- A copy of the UHBristol Proposal Form, Presentation Template, Report Template, Summary Form, and Action Form are available [online] via: http://www.uhbristol.nhs.uk/healthcare-professionals/clinical-audit/doing-projects-at-ubht.html
- The UHBristol Clinical Audit Central Office can be contacted on tel. (0117) 342 3614 or e-mail: stuart.metcalfe@uhbristol.nhs.uk
- Clinical Audit Training Workshops can be booked through the Clinical Audit Central Office.

CLINICAL EFFECTIVENESS

- For advice on Clinical Effectiveness, including how to write guidelines, contact James Osborne, Clinical Effectiveness Co-ordinator, tel. (0117) 342 3753 or e-mail: james.osbourne@uhbristol.nhs.uk
PATIENT ENGAGEMENT

- For advice on Patient Involvement, including designing structured surveys and questionnaires contact Paul Lewis, Patient Involvement Facilitator, tel. (0117) 342 3638 or e-mail: paul.lewis@UHBristol.nhs.uk
- For advice on Patient Involvement, including unstructured surveys and focus groups contact Tony Watkin, Public Involvement Lead, tel. (0117 342 3729 or e-mail: tony.watkin@UHBristol.nhs.uk
- Surveys MUST be approved by the Trust’s Questionnaire, Interview and Survey (QIS) Group. Proposals should be submitted to Paul Lewis using the QIS proposal form. The proposal form is available [online] via http://www.uhbristol.nhs.uk/healthcare-professionals/clinical-audit/doing-projects-at-ubht.html
- A copy of the UHBristol Covering Letter template is available [online] via the internal intranet site http://connect/Governance/patientexperience/ppi/Pages/QISGroup.aspx

RESEARCH

- For advice on research projects contact the Research & Development Department, tel. (0117) 342 0233 or e-mail: r&doffice@uhbristol.nhs.uk

LITERATURE REVIEWS

- For advice on literature reviews contact the Learning Resource Centre, tel. 0117 342 0105 or e-mail: learningresources@UHBristol.nhs.uk

SAMPLE SIZES