

How To: Analyse & Present Data

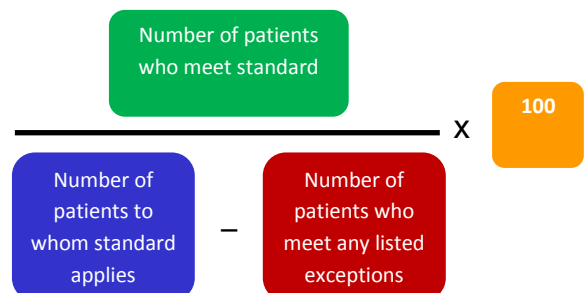
INTRODUCTION

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 free-text. 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 consider 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.

CALCULATING COMPLIANCE WITH CLINICAL AUDIT STANDARD



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. This is known as 'categorical' or 'nominal' data; data that can be sorted according to non-overlapping (mutually exclusive) categories, where each subject in a sample can only fit into one category. For example:

- Staff grade: Consultant, Registrar, Specialist Nurse
- Age group: 16-20, 21-25, 26-30, etc.
- Standard met: Yes / No

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, a variation of the percentage calculation is needed:

EXAMPLE 2:

- Audit criteria: All patients should attend a pre-op clinic
- Exception: emergency operation
- Results: Yes = 32, No = 5 and N/A (emergency) = 13

$$\frac{32}{50 - 13} = \frac{32}{37} \times 100$$

- 32 patients attended a pre-op clinic
- 18 did not, but 13 of these were emergencies (exceptions)
- **Therefore 32/37 (86%) met the standard**
- The difference between 64% in the first example, where no exceptions were taken into account, and 86% in this example is significant enough to influence our thinking about how well we are doing with meeting this standard, so it is important to remember your exceptions!

B. NUMERICAL DATA

Some of the data items you collect are likely to be numerical values, e.g.

- 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.

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:

Sum of all the values ÷ Number of values

The table to the right shows data about length of stay (LOS) on three wards.

Length of Stay (days)	Number of patients discharged		
	Ward 1	Ward 2	Ward 3
1	4	4	1
2	8	7	3
3	12	17	3
4	18	10	4
5	20	7	10
6	18	4	15
7	12	2	4
8	8	2	2
9	3	5	0
10	0	9	0

For Ward 1, the mean is:

$$\frac{(1 \times 4) + (2 \times 8) + (3 \times 12) + (4 \times 18) + (5 \times 20) + (6 \times 18) + (7 \times 12) + (8 \times 8) + (9 \times 3)}{4 + 8 + 12 + 18 + 20 + 18 + 12 + 8 + 3} = \frac{511}{103} = 4.96$$

The mean LOS on ward 1 is therefore 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 in Chart 1 below. The mean suggests that the data is the same for all three wards, however the chart indicates that this is not the case. The problem with the mean is what it does not tell us.

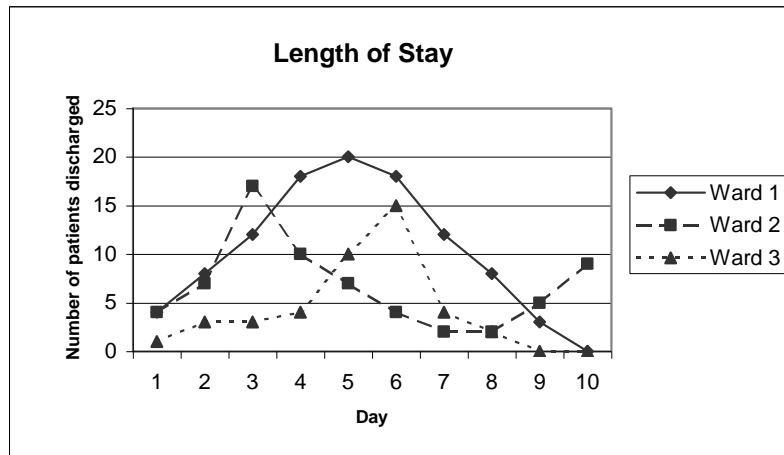


Chart 1

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 and obtaining figures for compliance with your standards is still the principal aim of analysing clinical audit data. 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 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 median), 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 categorical 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 (and could potentially therefore be added to the exception criteria for the standard in future) or will show what needs to be your 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 are close to 100%, it may be agreed that any further improvement will be difficult to obtain and that other standards with lower results should be 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%, while in other areas a lower (but still high) percentage might be considered acceptable.

2. DISPLAYING DATA

USING TABLES

The simplest way to present data is in a table and this is the best way to show compliance with your audit standards (see Table 1); in fact, our audit report template will ask you to provide this regardless of what else you do in the way of analysing and displaying your data.

Table 1: compliance with audit standard

Standard	Target	Result
All prescriptions should include frequency of dose	100%	80% (40/50)

Table 2: category of caesarean sections in audit

Category of urgency	Frequency	Percentage
Category 1	12	21%
Category 2	27	47%
Category 4 (elective)	18	32%

USING CHARTS

The purpose of a chart is the visualisation of data, i.e. providing useful information in a graphical form. It is simple enough to enter figures into a spreadsheet and hit the “insert chart” button, but in order for charts to be meaningful and useful to your audience a little more thought and preparation is often necessary.

The key principles to think about are:

- What message am I trying to convey?
- Do I need to represent the data graphically to get the message across?
- What kind of chart will deliver my message in the clearest way?
- What information could I include in order to anticipate and answer my audience’s questions?

- **Bar/column charts** – For categorical data – Generally used to show frequency, e.g. number of patients meeting the standard / not meeting the standard, or the number of patients seen by different staff groups.

For example, if your standard stated that all patients in A&E who meet certain criteria should be seen by a consultant, you might want to show your audience what grade of staff saw the patients if it was not the consultant (see Chart 2, right).



Chart 2

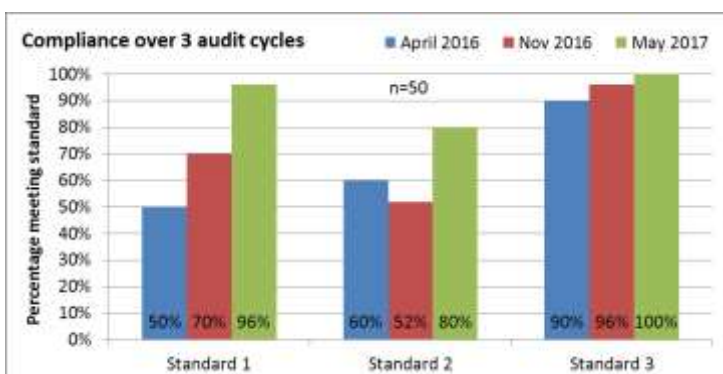


Chart 3

Versions of bar charts (stacked or comparative) – You can show more than one standard and/or more than one audit per chart, either by displaying bars with different values next to each other (Chart 3) or by stacking bars on top of each other (Chart 4).

Chart 4

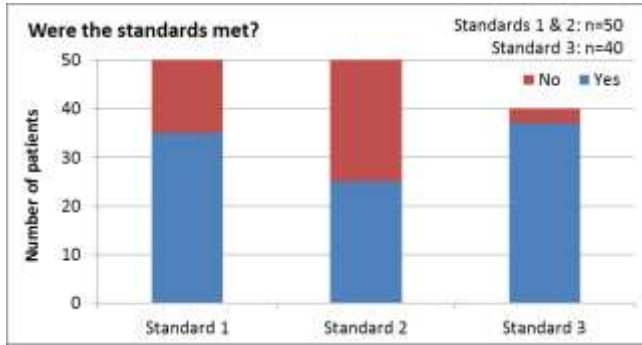
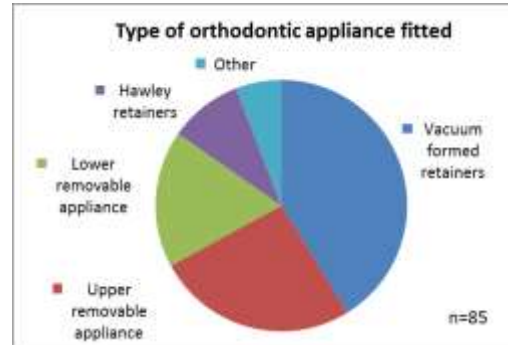


Chart 5

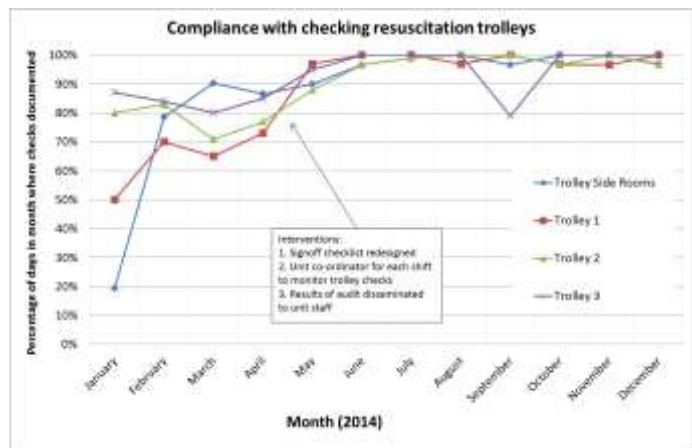


- **Pie charts** - For categorical data – Used to show proportion of parts compared to a whole, e.g. percentage compliance with a standard, or the types of device fitted in a sample of patients whose devices failed (see Chart 5). As a general rule, keep the number of slices low (advice differs on the maximum, but above 6 slices your chart may become hard to interpret) and avoid using this type of chart if the values are all very similar, as small differences in the size of slices are difficult to see.

Another occasion to avoid using a pie chart is when the parts do not add up to a meaningful whole; for example, if you wanted to illustrate ‘type of treatment’ in cases where patients are likely to have had more than one treatment each. 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 probably not very meaningful in itself. A bar chart could be used instead.

- **Line charts** – Can be used to show change over time, e.g. ongoing compliance with a regular monthly audit of hand hygiene or equipment checks

Chart 6

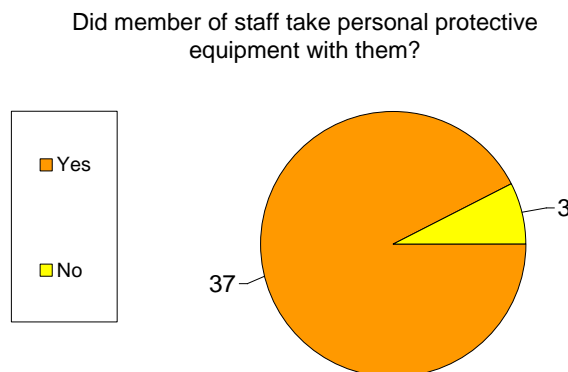


CREATING GOOD CHARTS

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.

EXAMPLE CHART 1:

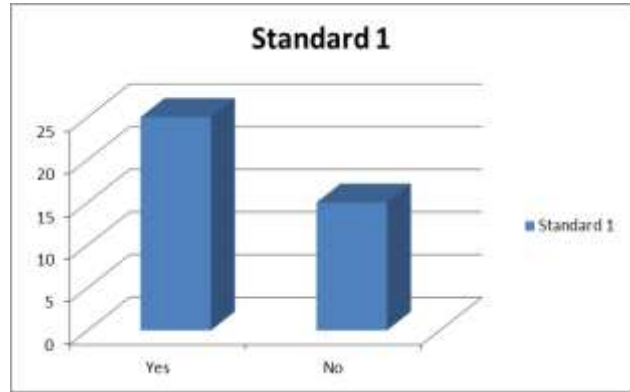
A chart might be considered unnecessary to illustrate this result. Consider whether people need to see a graphical representation of your data; in this case simply expressing the data as ‘37/40 members of staff (93%) took personal protective equipment on domiciliary visits’ should be sufficient. In a 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:

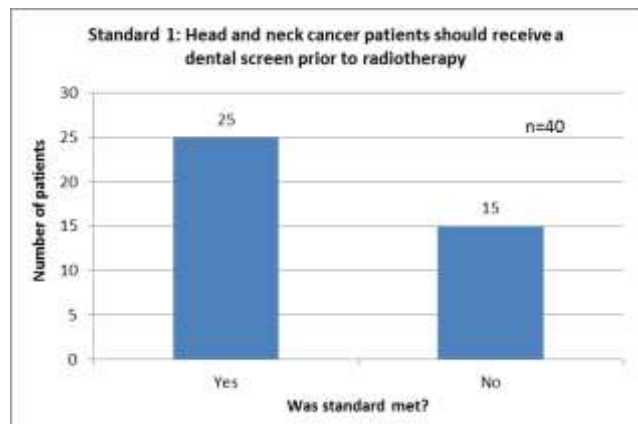
What's wrong?

- The 3D effect makes it difficult to read how many “yes” or “no” answers there are.
- 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 (unlike in the stacked or comparative bar charts shown in charts 3 and 4)
- There is a lot of white space on the page. It looks unprofessional.



Revised version

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



FURTHER READING

“An introduction to statistics for local clinical audit and improvement”, HQIP guide, 2015
<http://www.hqip.org.uk/resources/introduction-to-statistics-for-clinical-audit-and-qi>



CONTACT DETAILS/ USEFUL INFORMATION

CLINICAL AUDIT

- The UHBristol **Clinical Audit website** is available via <http://www.uhbristol.nhs.uk/for-clinicians/clinicalaudit/>
- Contact details for UHBristol **Clinical Audit Facilitators** are available via <http://www.uhbristol.nhs.uk/for-clinicians/clinicalaudit/contacts/>
- The full range of UHBristol Clinic Audit **'How To' guides** are available via <http://www.uhbristol.nhs.uk/for-clinicians/clinicalaudit/how-to-guides/>
- Copies of UHBristol **Clinical Audit Proposal Form, Presentation Template, Report Template, Summary Form, and Action Form** are available via <http://www.uhbristol.nhs.uk/for-clinicians/clinicalaudit/carrying-out-projects-at-uh-bristol/>
- The UHBristol **Clinical Audit & Effectiveness Central Office** can be contacted on 0117 342 3614 or e-mail: stuart.metcalfe@uhbristol.nhs.uk
- **Clinical Audit Training Workshops** can be booked through the Clinical Audit & Effectiveness Central Office as above.

CLINICAL EFFECTIVENESS

- For advice on **Clinical Effectiveness (NICE, NCEPOD, PROMS, guidelines)** matters contact Stuart Metcalfe, Clinical Audit & Effectiveness Manager, 0117 342 3614 or e-mail: stuart.metcalfe@uhbristol.nhs.uk

PATIENT EXPERIENCE

- For advice on carrying out **surveys, interviews and questionnaires** please contact Paul Lewis, Patient Experience Lead (Surveys & Evaluations), 0117 342 3638 or e-mail: paul.lewis@UHBristol.nhs.uk
- For advice on conducting **qualitative and Patient Public Involvement Activities (focus groups, community engagement, co-design, workshops)** please contact Tony Watkin, Patient Experience Lead (Engagement & Involvement), 0117 342 3729 or e-mail: tony.watkin@UHBristol.nhs.uk
- All surveys that are being carried out for service evaluation or audit purposes should be discussed with Paul Lewis in the first instance. Patient experience surveys will also usually need to 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 and covering letter template is available via <http://www.uhbristol.nhs.uk/for-clinicians/patient-surveys,-interviews-and-focus-groups/>

RESEARCH

- For advice on research projects contact the **Research & Innovation Department** on 0117 342 0233 or e-mail: research@UHBristol.nhs.uk
- Further information can be found via <http://www.uhbristol.nhs.uk/research-innovation/contact-us/>

LITERATURE REVIEWS/EVIDENCE

- For advice on literature reviews, NHS Evidence, article/book requests and critical appraisal contact the **Library and Information Service** on 0117 342 0105 or e-mail: Library@UHBristol.nhs.uk

SAMPLE SIZES

- The **Sample Size Calculator** is available via: <http://www.uhbristol.nhs.uk/for-clinicians/clinicalaudit/how-to-guides/>

QUALITY IMPROVEMENT

- Further information about clinical audit and wider quality improvement is available via the Healthcare Quality Improvement Partnership (HQIP) - <http://www.hqip.org.uk/>