2012 Accident and Emergency Survey
Redesign of the Benchmark Reports

The benchmark reports for the 2012 Accident and Emergency Survey are a new style of report and replace the previous reports produced for the national surveys which contained scores out of 100. The data contained here uses the same scoring system as before but presents the data as a score out of 10. However, the data has been analysed and therefore categorised differently, and displays trusts' performance in a different way to the previous reports, using a more robust statistical technique called the 'expected range', rather than identifying the top and bottom 20% of trust scores.

We have designed the reports using feedback from people who use the data, so that as well as meeting their needs, it presents the groupings of the trust results in a simpler and fairer way, to show where we are more confident that a score is 'better' or 'worse'.

The following is provided to answer some of the questions you may have on the changes that have been made, as well as on understanding and using your data. A technical guidance document is also available on the CQC website which goes into further detail on the statistical techniques used to categorise trust scores, and can be found here: **www.cqc.org.uk/accidentandemergency**

Questions and Answers

THE BENCHMARK REPORTS

Why have you made changes to the benchmark reports?

We received feedback that having two different ways of analysing and presenting the data is confusing to some, and it can be unclear which data to use. In the old style of benchmark reports, it was also necessary to take into account the confidence interval surrounding the score to accurately understand how a trust is performing. Feedback also showed that some people found these confusing and often the confidence intervals were not used.

Furthermore, the categories used in the old style of benchmark report were very simplistic and based on percentiles: ordering trust scores for each question by size and identifying the top 20% of trusts and bottom 20% of trusts. This meant that 29 trusts[1] would be in the top or bottom 20% irrespective of how well or how badly trusts may be performing on a particular question.

The advantages of the redeveloped benchmark reports are that:

- There are no longer two different ways of analysing and presenting the data
- There are no 'confidence intervals' surrounding the score to take into account when interpreting the data
- As the method of analysis is much more sophisticated, you can be **very confident** that if the score for your trust is showing in either the red of the green that your trust is performing better or worse than all other trusts.

Why are the scores presented out of ten rather than 100?

The scores are presented out of ten to emphasise that they are scores and not percentages. Please remember that the scores **are exactly the same** as trusts have been provided with previously but have simply been divided by ten. What has changed is the method of analysis and categorisation of the data in the benchmark reports.

[1] 147 trusts took part in the 2012 survey

What are the red, green and orange sections in the chart?

The red, green and orange sections in the chart display the expected range for a score for a trust. The orange section is the ‘expected range’, the green section shows where a score would lie if it were better than expected, and the red section signifies worse than expected performance. The expected range is where we would expect a particular trust to score if it performed about the same as most other trusts in the survey. The range takes into account the number of respondents from each trust as well as the scores for all other trusts, and allows us to identify which scores we can confidently say are 'better' or 'worse' than the majority of other trusts (for more detail please see the technical guidance for more details, available from: www.cqc.org.uk/accidentandemergency).

PLEASE NOTE: The expected range is **uniquely calculated for each trust for each question** and these groupings are **not the same** as those used in the previous style of benchmark report, which showed the top 20% and bottom 20% of scores. These groupings are instead based on a statistical analysis involving the use of adjusted Z scores and winsorisation. More detail can be found in the technical guidance, available from the link above.

Why are there no national thresholds?

We have moved away from this as it is a very simplistic way of analysing and presenting the data. Additionally, as set out in the above detailing why these changes were made, feedback suggested that having two different ways of analysing and presenting the data is confusing to some, and it can be unclear which data to use.

Why are there no confidence intervals surrounding the score?

As the ‘expected range’ calculation takes into account the number of respondents at each trust who answer a question, as well as the scores for all other trusts, it is not necessary to present confidence intervals around each score.

ABOUT THE ANALYSIS

How are the scores calculated?

Each NHS trust received a score out of 10 for each question that evaluates their performance, based on the responses given by their patients. A higher score is better. The scoring system is exactly the same as has been used in previous years, just out of ten rather than 100 and is described in the technical guidance available from the CQC website.

How is the data analysed?

A technical document describes in detail how the trust level data is scored, standardised and analysed. This is published on the CQC website and available from the surveys team.

In summary, the trust level data is calculated by converting responses to particular questions into scores. These were calculated by converting each respondent’s answer to a question into a score (from 0 to 10) then averaging these to arrive at a single score for the trust, for each question. The higher the score, the better a trust is performing. The scored data has been ‘standardised’ by the age and gender of respondents.

The calculation used to analyse the trust level data is a statistic called the ‘expected range’. This means that a lay person does not need to interpret any statistical details as this interpretation has been carried out for them. It is a robust analysis as it takes the reliability of the data into account and we can be extremely confident that a trust is performing ‘better’ or ‘worse’ than average using this analysis.

Why is the data standardised / weighted by the age and gender of respondents?

The reason for ‘standardising’ (or weighting) data is that we know that the views of a respondent can reflect not only their experience of NHS services, but can also relate to certain demographic characteristics, such as their age and sex. For example, older respondents tend to report more positive experiences than younger respondents, and women tend to report less positive experiences than do men. Because the mix of patients varies across trusts (for example, one trust may serve a considerably older population than another), this could potentially lead to the results for a trust appearing better or worse than they would if they had a slightly different profile of patients. To account for this we ‘standardise’ the data. Standardising data adjusts for these differences and enables the results for trusts to be compared more fairly than could be achieved using non-standardised data.

What is the ‘expected range’?

The better / about the same / worse categories are based on the 'expected’ range that is calculated for each question for each trust. This is the range within which we would expect a particular trust to score if it performed about the same as most other trusts in the survey. The range takes into account the number of respondents from each trust as well as the distribution of scores for all other trusts, and allows us to identify which scores we can confidently say are 'better' or 'worse' than the majority of other trusts (see Appendix B for more details). Analysing the survey information in such a way allows for fairer conclusions to be made in terms of each trust’s performance. This approach presents the findings in a way that takes account of all necessary factors, yet is presented in a simple manner.

It is the same analysis technique as applied to the risk ratings in the Quality and Risk Profiles, and is based on identifying outliers through the use of adjusted Z scores. More detail on this is available in the technical guide linked to from each trusts results, and available from the surveys team.

Why do the results and / or number of respondents provided by CQC differ from those provided to me by our approved contractor?

CQC do not see the reports provided to you by your approved contractor and therefore cannot comment on these. You should raise any queries directly with your approved contractor. However, likely reasons for any discrepancies are:

- The approved contractor may have cleaned the data differently to CQC. In particular, CQC remove respondents from the base of a question that do not analyse the performance of a trust - we refer to these as ‘non specific responses’, such as ‘don’t know or can’t remember’. A guide to data cleaning is available at: **www.nhssurveys.org/surveys/639**
- Trust level data published by CQC has been ‘standardised’ or ‘weighted’ for age and gender to enable fairer comparisons between trust results. Approved Contractors may not have done this or may have applied a different standardisation. To be able to standardise the data, information is needed on both

age and gender, if either of these pieces of information is missing the respondent must be dropped from the analysis as it is not possible to apply a weight.

- CQC analyses trust level data by scoring (and standardising) the responses to each question. Each response option that evaluates performance is scored on a scale of 0-10. Approved Contractors may have analysed and / or scored the data in a different way.
- The Approved Contractor will not be able to make comparisons against all trusts that took part in the survey, only against those that commissioned them. Therefore any 'national' results they publish will not be based on all trusts and any thresholds they calculate may be different.

UNDERSTANDING THE DATA

The score for one of my questions has gone up but is categorised as 'about the same' yet last year we were 'better'?

When looking at scores within a trust over time, it is important to be aware that they are relative to the performance of other trusts. If, for example, a trust was 'better' for one question, then 'about the same' the following year, it may not indicate an actual decrease in the performance of the trust, but instead may be due to an improvement in many other trusts' scores, leaving the trust to appear more 'average'. Hence it is more useful to look at actual changes in scores and to test for statistically significant differences.

We are categorised as 'about the same' for a question yet a trust with a slightly lower score than us is categorised as 'better'. Why is this?

The 'expected range' calculation takes into account the number of respondents from each trust as well as the distribution of scores for all other trusts, and allows us to identify which scores we can confidently say are 'better' or 'worse' than the majority of other trusts. As set out above the expected range is a conservative statistic: it accounts for the possibility that there is variation across trusts for other reasons, aside from differences in trust performance. There may be significant variation between trusts due to certain factors that are not within the trusts' control. The technique used takes this into account. It is likely that your trust came out as 'about the same' because your trust had fewer respondents to the question which creates a greater degree of uncertainty around the result. The trust with the lower score would likely have had more respondents to the question, and so their expected range would have been narrower.

Why do most trusts appear to be performing 'about the same'?

The expected range is a conservative statistic. It accounts for the possibility that there is variation across trusts for other reasons, aside from differences in trust performance. There may be significant variation between trusts due to certain factors that are not within the trusts' control. The technique used takes this into account, and so if a trust is found to be performing 'better' or 'worse' compared with most other trusts that took part in the survey, you can be really very confident that this is the case and it is extremely unlikely to have occurred by chance.

Even though your trust may appear to be performing 'about the same' compared to most other trusts nationally, the results should still be useful to you locally, for example you may want to:

- Make comparisons to the results from previous surveys to look for questions where you have improved or declined.
- Identify particular areas you may wish to improve on ahead of the next survey

- Compare your results with those of other similar trusts.
- Look at your results by different type of patient to understand their different experiences, for example, by age, ethnic group or day of attendance.
- Undertake follow up activity with patients such as interviews, workshops or focus groups to get more in depth information into areas in which you would like to improve.

Please remember that for points 1-3 above, to do this accurately you should undertake an appropriate significance test. Please see the below FAQ which provides advice on making comparisons between different survey years.

Please see chapters 18, 19 and 20 of the survey guidance manual for more information on using survey data. The guidance manual is available on the NHS surveys website, please see the further information section.

How do I calculate an overall score for my trust?

It is also important to remember that there is no overall indicator or figure for 'patient experience', so it is not accurate to say that a trust is the 'best in the country' or 'best in the region' *overall*. Adding up the number of 'better' and 'worse' categories to find out which trust did better or worse overall is misleading: we do not provide a single overall rating for each NHS trust as this would be too simplistic. The survey assesses a number of different aspects of patient experience (such as health and social care staff, medication, care plans etc) and trust performance varies across these different aspects. This means that it is not possible to compare the trusts overall. It is better to look at the trusts in your area and see how they perform across the particular aspects that are most important to you.

COMPARING RESULTS

Why is statistical significance relevant?

A statistically significant difference means that the change in the results is very unlikely to have occurred by chance. Without significance testing you cannot be sure that a difference between two results would still be different if you repeated the survey again. If a result is not significant then you cannot be sure of its accuracy. If a significant difference is present then it is likely that it is a true difference, and if the survey was repeated again that you would see the same outcome.

Which trusts are performing best / worst?

We have compiled a list of all trusts that performed better or worse when comparing data across all trusts, for each scored question in the survey which is available from the surveys team on request upon publication. This can be used to at a glance identify which trusts are in each group, rather than searching through each individual trust page or benchmark report. Please note the 'interpretation' information at the beginning of the document, which explains how the information should be most appropriately reported.

Why does the number of trusts performing 'better' or 'worse' at each question vary?

It is important to be aware that the ranges of performance on different questions varies and this has an influence on how much a trust needs to differ from the average by, in order to be considered 'better' or 'worse' than the average. This means that the number of trusts to perform 'better' or 'worse' at each question will vary.

Why has no trust come out as performing better or worse for a particular question?

This can occur in the analysis of the data and is an acceptable consequence of the statistical technique that is used. The size of the expected range is constructed by considering how different all trust scores are across the range, as well as the confidence we can have in that particular trust's score (by looking at the number of respondents to that question). In some cases, this will lead to such a wide margin of error that the 'expected range' will be very wide, and hence will also cover the highest or lowest scoring trusts for that question.

Is the lowest scoring trust the worst trust in the country, for each question? And likewise the highest scoring trust the best?

If a trust is in the 'better' or 'worst' category this mean that they are performing either better or worse compared with *most other trusts* that took part in the survey. However, a trust is not necessarily the best, or the worst, and this could not be determined without undertaking an appropriate significance test.

If you took the scores and ordered them by size, you would most likely find that the highest and lowest ones would change if you ran the survey again. This is because the scores are estimates – we have only had questionnaires from some patients who were seen by the trust during the sampling period, not all patients. If another sample of patients were surveyed, and you put the scores in order again, you would find that there would probably be a different trust at the top and at the bottom. By analysing the data the way we have, we can say which trusts are likely to always be above average and those that will always be below average, so they should be looked at as a group, rather than in order of scores. This is the fairest way to present the data as it means that individual trusts are not pulled out as the very 'best' or very 'worst', when that may not be the case and it may be that if all patients were surveyed, different trusts would be shown to be the very 'best' or 'worst'.

How can I make comparisons to previous years data?

At the end of the benchmark report for your trust, in the table containing the data, you will find the results for your trust from the 2008 survey. The final column is called 'change from 2008' and displays arrows to indicate whether the score for this years survey shows a statistically significant increase (up arrow), a statistically significant decrease (down arrow) or has shown no statistically significant change (no arrow), when compared with 2008. A statistically significant difference means that the change in the results is very unlikely to have occurred by chance. Where a result for 2008 is not shown, this is because the question was either new this year, or has had the question wording and / or the response categories changed, meaning that it is not possible to compare the results as we do not know if any change is caused by alterations in the survey instrument, or variation in service performance.

Please note that comparative data is only shown for question data, and not for the section data. We review the questionnaire every year to consider if any improvements are needed. If a change is made to a question and / or its response category, or a question is removed and / or a new question added, the data for the corresponding section into which the question falls will no longer be comparable. Therefore maintaining year on year comparability of section data is more difficult. We would also encourage trusts to look at their detailed question level results, as a summary score can hide particular areas where trusts may be doing well, or not so well.

Results were tested for significance using two-sample t-tests. If you want to make similar comparisons to the 2008 data, you would need to undertake the same significance test. Please be careful to ensure that you make accurate comparisons:

- Do not compare questions that are not comparable due to changes being made to the question wording, and / or the response categories. Questions that CQC considers to be comparable between 2008 and 2012 are shown in the national tables for 2012 which may be found at: **www.cqc.org.uk/accidentandemergency**

Why can't I sort the scores to rank the trusts in order of performance?

It is not appropriate to sort the scores:

1) Firstly, due to the analysis technique applied, where the number of respondents is taken into account, it is possible that one trust may score higher than another - though the higher scoring trust is classed as 'about the same' and the second, lower scoring, trust is put into the 'better' category. This may occur if the second trust has a considerably larger number of respondents, as it will be assumed that their score is more reliable, and hence more likely always to be high.

2) Secondly, the statistical technique does not measure how different individual trust scores are from one another (whether statistically significant), and so it would be too simple to attempt to sort by scores alone, without running more analysis on the data. The banding technique used is helpful in identifying which trusts are likely always to be in the 'better', 'worse', or 'about the same' category, no matter how many surveys are sent out.

FURTHER INFORMATION

Further Questions

If you have any further questions please contact the surveys team at CQC:
patient.survey@cqc.org.uk

The full national results for the 2012 survey are on the CQC website, together with an A to Z list to view the results for each trust (alongside the technical document outlining the methodology and the scoring applied to each question):
www.cqc.org.uk/accidentandemergency

The results for previous A&E surveys can be found on the NHS surveys website at:
www.nhssurveys.org/surveys/296

Full details of the methodology of the survey can be found at:
www.nhssurveys.org/survey/1145

More information on the programme of NHS patient surveys is available at:
www.cqc.org.uk/public/reports-surveys-and-reviews/surveys

More information on Quality and Risk Profiles (QRP) can be found at:
www.cqc.org.uk/organisations-we-regulate/registered-services/quality-and-risk-profiles-qrps

CQC Surveys team
November 2012