2013/2014 Ambulance 'Hear and Treat' Survey
Benchmark Reports: Q&A

This document is provided to answer some of the questions you may have on the benchmark reports, and on the underlying 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:
http://www.cqc.org.uk/Ambulancesurvey201314

Questions and Answers

The Benchmark Reports

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

The coloured bars represent the full range of all trust scores, from the lowest score achieved by a trust to the highest. The orange section in the charts represents the **expected range** for a score for a 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. If a score falls above or below the expected range it will be in the 'better' or 'worse' category, represented by green and red areas respectively. These categories are listed as 'above', 'average', and 'below' in the data shown on the CQC website. The calculation of the expected 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 (see the technical guidance for more details, available from: **http://www.cqc.org.uk/Ambulancesurvey201314** and sent to survey trust leads prior to publication).

How do I know which category my trust's score is in if the diamond representing the score appears to be on the threshold in the benchmark charts?

Text to the right of the graphs clearly states if a trust score for a particular question, or section, is 'better' or 'worse' compared with most other trusts that took part in the survey. If there is no text present, the result is 'about the same'.

How do I refer to these scores and categories when reporting on the results for my trust?

We have produced a brief guide on how to refer to the findings when disseminating the scored data. This was provided to survey leads prior to publication, and is available on request from the surveys team, please contact: patient.survey@cqc.org.uk.

About the Scores

Why are the scores presented out of ten?

The scores are presented out of ten to emphasise that they are scores and not percentages of respondents.

How are the scores calculated?

For each question in the survey, the **(standardised)** individual responses are converted into scores on a scale of 0 to 10. A score of 10 represents the best possible response and a score of zero the worst. The higher the score for each question, the better the trust is performing. For more detailed information on the methodology, including the scores assigned to each question, please see the technical document.

About the Analysis

What is the 'expected range'?

The better / about the same / worse categories are based on a statistic called 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. 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.

Why is the data standardised by the age and sex of respondents?

The reason for 'standardising' 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 with different population profiles to be compared more fairly than could be achieved using non-standardised data.

Why are the percentage results for all trusts not provided?

The percentage data is provided to trusts for their own information only as it can only be used to understand the results for individual trusts.

It is not suitable to use that data to make comparisons between trusts because the results are not **standardised** meaning that differences in the profiles of respondents are not taken into account. Any differences across trusts that are shown in non-standardised data may be in part due to differences in the characteristics of respondents. We know that age and sex are two such characteristics and so we adjust for this in the data to make fairer comparisons across trusts with differing population profiles.

A further advantage of using scored data is that it allows for all response options to be taken into account, rather than looking at just a subset of responses from the question. For example, if you look at the table below, from looking at the 'yes definitely' responses only, you would think that trust A and trust B are performing similarly. However, taking into account the other responses, it may be seen that trust B has the more positive result.

ReassureCT: Was the [first] person you spoke to reassuring?

	Trust A	**Trust B**
Yes completely	59%	59%
Yes to some extent	10%	39%
No	31%	2%

Scored, standardised data is therefore considered to be the fairest way to include survey data in the Commission's regulatory activities, as well as by NHS England for their measures and assessments.

In the past the percentage results or scores have been used to present data in a league table form, or to identify the ‘best’ or ‘worse’ trusts. Such use would be misleading and inaccurate, as the differences have not been tested for significance.

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.

Understanding the Data

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:

- Identify particular areas you may wish to improve on.
- Compare your results with those of other similar trusts.
- Look at your results by different service user groups to understand their different experiences, for example, by age, gender, ethnic group, etc.
- 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.
- Consider contacting other trusts doing well in areas where you want to improve to find out more about what they may be doing.

Please remember that for points 2 and 3 above, to do this accurately you should undertake an appropriate **significance test**.

The survey guidance manual provides more information on making use of survey data. The guidance manual is available on the NHS surveys website, please see the further information section. If using an approved contractor, they should also be able to provide you with advice.

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 most other trusts by, in order to be considered ‘better’ or ‘worse’ than most other trusts. This means that the number of trusts to perform ‘better’ or ‘worse’ at each question will vary.

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 wors*t, 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 responses from some callers who were selected during the sampling period, not all callers. If another sample of callers 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’.

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 is one of my sections labelled as ‘worse’ yet all of the questions that fall into that section are ‘about the same’?

If this has occurred, it is likely that your trust scored very lowly or even on the threshold for all or most of the questions that are in a section.

The thresholds for ‘worse,’ ’about the same’ and ‘better’ are based on the score variance. For sections, this is a composite of the separate question variances, but not a straightforward sum, because it also depends on the correlation between questions. It does not therefore follow that a trust that is above the threshold on separate questions will also be above the threshold when those questions are combined.

The ‘expected range’ is dependent on the (sampling) variance of the trust’s results – with a more reliable score (as would normally be the case for section scores due to the pooled data), it is easier to be significantly different from the ‘average’ group than for a less reliable score.

How do I calculate an overall score for my trust?

It is 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 throughout the duration of receiving advice over the telephone, 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 that are similar to yours, or particular trusts against which you want to compare yourself, and see how they perform across the particular aspects that are of interest to you.

Comparing Results

Why is statistical significance relevant?

Survey scores are estimates – we have only received responses from some callers during the sampling period, not all callers, as some choose not to respond or could not be contacted. If another sample of callers were surveyed, you may find the results would change slightly. This is why it is important to test results for statistical significance.

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.

How can I make comparisons to other trusts?

The purpose of the expected range is to arrive at a judgement of a how a trust is performing compared with all other trusts that took part in the survey. To use the data in another way, by making comparisons between trusts, you will need to undertake an appropriate statistical test to ensure that any difference is statistically significant. A statistically significant change means that you can be very confident that the difference is real and not due to chance.

For advice on making accurate comparisons you may like to speak to someone within your trust with statistical expertise. The guidance documents issued with previous benchmark reports for other national surveys included some advice on using confidence intervals to check for statistically significant differences across scores, see for example section four in the following document:
http://www.nhssurveys.org/Filestore/documents/IP10_Guide_to_benchmark_reports.pdf

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 can't I sort the scores for all trusts and 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 analyses 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 carried out.

Further information

The full national results are on the CQC website, together with an A to Z list to view the results for each ambulance trust alongside the technical document outlining the methodology and the scoring applied to each question. Please note, the categories on the website are labelled 'above', 'average' and 'below', rather than the 'better', 'worse' and 'about the same' as used in the benchmark reports:
www.cqc.org.uk/Ambulancesurvey201314

Full details of the methodology of the survey can be found at:
www.nhssurveys.org/surveys/285

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 CQC's role in regulating, checking and inspecting services is available on the CQC website:
www.cqc.org.uk/

Further Questions

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

CQC Surveys team
June 2014