2016 Emergency Department Survey
Benchmark Data: FAQ

This document is provided to answer some of the questions you may have on the trust level results, as provided in the benchmark reports, and on the CQC website. A technical document is also available on the CQC website which goes into further detail on the statistical techniques used to categorise trust scores, and a Quality and Methodology outlines further details on the survey.

The Benchmark Reports

Why are no benchmarked results available for Type 3 departments?

In previous years, the emergency department survey focussed solely on Type 1 departments. For 2016, to reflect recent changes in the provision of urgent and emergency care, the survey was expanded to also include people who attended Type 3 departments that are provided directly by the acute trust. If a trust did not have a Type 3 department, their sample was drawn from their Type 1 service only.

One hundred and thirty seven trusts took part in the survey, and of these, 49 trusts had both a Type 1 and a Type 3 emergency department and 88 trusts had only a Type 1 emergency department.

Benchmarked results have not been provided for Type 3 departments for the following reasons:

- Only Type 3 departments run by acute trusts were included, and not those run in those run in collaboration with, or exclusively by others. This means that we only have a partial picture of peoples' experiences of Type 3 departments in England;
- For trusts that had a Type 1 and Type 3 department, 950 people who used Type 1 services were sampled and 300 people who used Type 3 services were sampled. Therefore the data set for Type 3 patients is quite small. The Type 3 patient data can be insightful when looking at the results derived from the pooled data for all 49 trusts that had a Type 3 department. However, when looking at this data at trust-level, there are a very large numbers of missing values, and this combined with the lower sample size means the chances of sampling error (the chance that the sample is not representative of the actual population) are high. Type 3 data should only be used at the trust-level with extreme caution.

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. 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 on the CQC website, and sent to trust survey 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 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 trust survey leads prior to publication, and is available on request from the surveys team at: patient.survey@cqc.org.uk

About the Scores

Why are the percentage results for all trusts not provided?

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

It is not suitable for making 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 gender 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 becomes apparent that trust B has the more positive result overall.

Q12: Did you have enough time to discuss your health or medical problem with the doctor or nurse?

	Trust A	**Trust B**
Yes definitely	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 other stakeholders such as NHS England and the Department of Health 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 'better' or 'worse' trusts. Such use would be misleading and inaccurate, as the differences have not been tested for significance.

Why are the scores presented out of ten?

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

How are the scores calculated?

For each question in the survey that can be scored, 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 example, Q12 is scored as per the example below. The option of 'No' was allocated a score of 0, as this suggests that the respondents' experiences need to be improved. A score of 10 was assigned to the option 'Yes, definitely', as it reflects a positive experience. The remaining option, 'Yes, to some extent', was assigned a score of 5 as respondent did not feel fully listened to. Hence it was placed on the midpoint of the scale. The 'I did not see a doctor or nurse' response option is not scored as this does not evaluate trusts in any way.

Q12: Did you have enough time to discuss your health or medical problem with the doctor or nurse?	
Yes, definitely	10
Yes, to some extent	5
No	0
I did not see a doctor or nurse	Not applicable

For more detailed information on the methodology, including the scores assigned to each question, please see the technical document and / or the scored questionnaire.

Why aren't all questions scored?

Not all questions are appropriate for scoring as they do not evaluate trust performance. For example, they may be 'routing questions' designed to filter out respondents for whom the following questions do not apply. For example, Q4 (Were you taken to the emergency department in an ambulance?) is a routing question, as those who were not are instructed not to answer Q5. Other questions that can't be scored are descriptive, for example the demographic questions such as Q47 (Are you male or female?).

About the Analysis

What is the expected range?

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

If a trust is categorised as 'better' or 'worse' than average then we can be **very confident** that it would continue to appear better or worse than average if the survey was repeated with a different sample

More detail on this is available in the technical document.

Why is the data standardised by the age and gender 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 gender. 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 people who use services 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 people who use services. 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 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 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 in England, the results should still be useful to you locally, for example you may want to:

- Compare your results with those of other similar trusts.
- Look at your results by different respondent groups to understand their different experiences, for example, by age, gender, ethnic group, site/location etc.
- Identify particular areas you may wish to improve on ahead of future surveys.
- Undertake follow up activity with people who use your services such as interviews, workshops or focus groups to get more in depth information into areas in which you would like to improve.
- Review the feedback provided in the 'other comments' section of the questionnaire. If you are using an approved contractor, they may analyse this for you, depending on what you have agreed with them.

Please remember that for the first two points above, to do this accurately you should undertake an appropriate significance test.

More information on making use of survey data is available on the NHS surveys website.

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 performing '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 and / or lowest scoring trusts for that question. This could be because there were few respondents and / or there was a great deal of variation in their answers.

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 means 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 test for statistical significance.

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 people who used services during the sampling period (September 2016), not everyone. If another sample of people who use services 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 'better' and those which are likely to always be 'worse', so they should be looked at as a group of 'better' trusts and 'worse' trusts, 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 everyone who used services 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 the badge category for one of my sections 'worse' yet all of the questions that fall into that section are 'about the same'?

This can happen because the calculation of the section scores is a separate calculation and not an average of all questions that make up a particular section. 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) 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 ‘service user 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 people’s experiences and trust performance varies across these different aspects. This means that it is not recommended 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.

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’ and include options such as such as ‘don’t know or can’t remember’. For more information please see the data cleaning guide.
- Trust level data published by CQC has been ‘standardised’ by age and gender to enable fairer comparisons between the results of trusts which may have different population profiles. 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, or not able to be determined, 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 overall results they publish will not be based on all trusts and any thresholds they calculate may be different.

Comparing Results

Why is statistical significance relevant?

Survey scores are estimates – we have only received questionnaires from some people who used services during the sampling period (September 2016), not everyone, as the survey uses a sample of 1250 and of these, some choose not to respond. If another sample of people 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 results is very unlikely to have occurred by chance. Without significance testing you cannot be sure that a difference between two results would still occur 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?

Open data is published on the CQC website which contains the results for all trusts.

However, when using this data, it is important to note that with the analysis technique used, all we can say is that a particular trust is 'significantly worse' or 'significantly better' than most other trusts.

We cannot say, for example, that that a trust that has a score of 4.5 (Trust A) is any better than a trust with a score of 4.3 (Trust B). To do so we would need to carry out a statistical test to determine whether this difference is statistically significant. If a difference is not significant, to say one trust is better than another is unfair and inaccurate.

We have also published a separate report identifying outlier trusts and this is available on the CQC website. Due to the smaller number of responses for Type 3 departments, this analysis is available for Type 1 departments only.

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: to make comparisons to other trusts, you will need to undertake an appropriate statistical test to ensure that any difference in scores is statistically significant. A statistically significant difference 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, or your approved contractor (if used) should be able to advise on this, or undertake this analysis for you depending on your contractual arrangements. The guidance documents issued with previous benchmark reports included some advice on using confidence intervals to check for statistically significant differences across scores.

Why are results from the 2016 survey not comparable to those from previous surveys?

There have been a number of changes to the sampling and analysis methodology for the 2016 survey, which means it is **not comparable** with other surveys (carried out in 2003, 2004/5, 2008, 2012 and 2014) for the following reasons:

- The sample size was increased from 850 to1,250
- The sample month for the 2016 survey (September) has changed from the 2014 survey (January, February or March)
- Previous surveys have focussed on major A&E Departments (Type 1) only. This survey also includes Type 3 Departments run directly by the acute trust
- The method used to weight the results has changed. For more information please see the Quality and Methodology report
- The provision of urgent and emergency services has changed over recent years, and is still evolving. This means that it would not be fair to compare the results from the 2016 survey to earlier surveys, as the landscape has changed.

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 by size for two reasons:

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 to remain high if the survey were repeated.

2) Secondly, the statistical technique does not measure how different individual trust scores are from one another (i.e. whether statistically significant), and so it would be too simplistic 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.

Can I see results for my local clinic / site etc.?

The survey data is currently presented at trust level only. At present we are unable to provide data at a more granular level for two reasons. Some sites may have too few respondents to achieve sufficient numbers of respondents (we set the cut off limit as 30 respondents per organisation). Given that the survey is used by other stakeholders such as NHS England, the Department of Health and others to measure trends over time, we are currently unable to change the sampling to accommodate this, without affecting the comparability across years. However, trusts are able to increase their sample size to enable this at a local level. Advice on how to do this was provided in the survey instruction manual.

Further information

The results for England and trust level results can be found on the CQC website. You can also find a technical document here, which describes the methodology for analysing the trust level results and a Quality and Methodology report:
www.cqc.org.uk/emergencydepartmentsurvey

The trust results from previous emergency department surveys are available at the link below. However, please note that results from the 2016 survey are **not comparable** with previous surveys. For more information on this, please see the statistical release or the Quality and Methodology report:
www.nhssurveys.org/surveys/296

Full details of the methodology for the survey, including questionnaires, letters sent to patients, instructions on how to carry out the survey and the survey development report, are available at:
www.nhssurveys.org/surveys/957

More information on the patient survey programme, including results from other surveys and a programme of current and forthcoming surveys can be found at:
www.cqc.org.uk/content/surveys

More information about how CQC monitors hospitals is available on the CQC website at:
www.cqc.org.uk/content/monitoring-nhs-acute-hospitals

Further Questions

We welcome all feedback and questions. Please contact the surveys team at CQC:
patient.survey@cqc.org.uk

CQC Surveys team
September 2017