

Care Quality Commission (CQC)

**Technical details
2016 Emergency Department Survey
October 2016**

Contents

1. Introduction	1
2. Selecting data for the reporting.....	1
3. The CQC organisation search tool	2
4. The trust benchmark reports	2
5. Interpreting the data	2
6. Further information	4
Appendix A: Scoring for the 2016 emergency department survey	5
Appendix B: Calculating the trust score and category	13
Appendix C: Calculation of standard errors	19

1. Introduction

This document outlines the methods used by the Care Quality Commission to score and analyse the trust level results for the 2016 Emergency Department Survey, as available on the Care Quality Commission website, and in the benchmark reports for each trust.

The survey involved 137 acute and specialist NHS trusts with a Type 1 emergency department¹. Forty nine of these trusts also had direct responsibility for running a Type 3 department² and patients from these departments were included within the survey for the first time in 2016, to reflect recent changes in the provision of urgent and emergency care. However, we have only benchmarked the data for type 1 departments.³ Please see the [Quality & Methodology report](#) for further information on why benchmark results for Type 3 departments have not been produced.

The survey results are available for each trust on the [CQC website](#). The survey data is shown in a simplified way, identifying whether a trust performed 'better' or 'worse' or 'about the same' as the majority of other trusts for each question. This analysis is done using a statistic called the **expected range** (see section 5.3). On publication of the survey, an A-to-Z list of trust names will be available.

The CQC webpage also contains a statistical release document containing England level results, alongside relevant national policy. Further information on the survey is available in the [Quality & Methodology report](#).

A benchmark report is also available for each participating trusts. Results displayed in the benchmark report are a graphical representation of the results displayed for the public on the CQC website (see further information section 6). These have been provided to all trusts and will be available on the [survey co-ordination centre website](#).

2. Selecting data for the reporting

Scores are assigned to responses to questions that are of an evaluative nature: in other words, those questions where results can be used to assess the performance of a trust (see section 5.1 for more detail). Questions that are not presented in this way tend to be those included solely for 'routing' respondents past any questions that may not be relevant to them (such as: 'In the last 12 months, have you been receiving any medicines for your mental health needs?') or those used for descriptive or information purposes (such as: Q25: Did you have any tests (such as x-rays, scans or blood tests) when you visited the emergency department?).

The scores for each question are grouped on the website, and in the benchmark reports for each trust, according to the sections of the questionnaire as completed by respondents.

¹ A Type 1 department is a major, consultant led A&E Department with full resuscitation facilities operating 24 hours a day, 7 days a week.

² A Type 3 department is an A&E/minor injury unit with designated accommodation for the reception of accident and emergency patients. The department may be doctor or nurse-led, treats at least minor injuries and illnesses and can be routinely accessed without appointment.

³ As the only trust with a Type 2 Emergency Department, Moorfields Eye Hospital NHS Foundation Trust has been treated as a Type 1 department within the benchmarking analysis.

Alongside both the question and section scores on the website are one of three statements:

- Better
- About the same
- Worse

This analysis is done using a statistic called the **expected range** (see section 5.3)

3. The CQC organisation search tool

The organisation search tool contains information from various areas within the Care Quality Commission's functions. The survey data for the emergency department survey can be found from the **A-Z link**. The presentation of the survey data was designed 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 simple and fair way, to show where we are more confident that a trust's score is 'better' or 'worse' than we'd expect, when compared with most other trusts.

Or by searching for a provider from the [CQC home page](#) then clicking on 'Surveys'.

4. The trust benchmark reports

Benchmark reports should be used by NHS trusts to identify how they are performing in relation to the other trusts who participated. The reports are available from the [Survey Co-ordination Centre website](#).

The graphs included in the reports display the scores for a trust, compared with the full range of results from all trusts that took part in the survey. Each bar represents the range of results for each question across all trusts that participated in the survey. In the graphs, the bar is divided into three sections:

- If a trust score lies in the orange section of the graph, the trust result is 'about the same' as most other trusts in the survey
- If a trust scores lies in the red section of the graph, the trust result is 'worse' than expected when compared with most other trusts in the survey.
- If a score lies in the green section of the graph, the trust result is 'better' than expected when compared with most other trusts in the survey

A black diamond represents the score for this trust. The black diamond (score) is not shown for questions answered by fewer than 30 people because the uncertainty around the result would be too great.

5. Interpreting the data

5.1 Scoring

Questions are scored on a scale from 0 to 10. Details of the scoring for this survey are available in Appendix A at the end of this document.

The scores represent the extent to which the respondent's experience could be improved. A score of 0 was assigned to all responses that reflect considerable scope for improvement, whereas a response that was assigned a score of 10 referred to the most positive experience reported. Where a number of options lay between the

negative and positive responses, they were placed at equal intervals along the scale. Where options were provided that did not have any bearing on the trust's performance in terms of peoples' experience, the responses were classified as "not applicable" and a score was not given. Where respondents stated they could not remember or did not know the answer to a question, a score was not given. The average score for all respondents within each trust is then presented, having applied standardisation (see 5.2 below).

5.2 Standardisation

Results are based on standardised data. 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 men. Because the mix of people using 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 of people using services. 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. In most cases this will not have a large impact on trust results; it does, however, make comparisons between trusts as fair as possible.

The Emergency Department Survey is standardised by **age and gender**.

5.3 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 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 C 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.

As the 'expected range' calculation takes into account the number of respondents at each trust who answer a question, it is not necessary to present confidence intervals around each score for the purposes of comparing across all trusts.

5.4 Conclusions made on performance

It should be noted that the data only shows performance relative to other trusts: we have not set out absolute thresholds for 'good' or 'bad' performance. Thus, a trust may score lowly relative to others on a certain question whilst still performing very well on the whole. This is particularly true on questions where the majority of trusts score very highly.

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 will be misleading if the limitations are not

recognised. The number of questions on each aspect of patient experience in the survey varies, and so will trusts' performance across these. So if you counted across all of the questions, some topics will have more influence on the overall average than others, when in fact some might not be so important. For more detailed scrutiny of performance, it is more comprehensive to look at trusts individual reports and assess performance by individual questions.

5.5 Comparing scores across or within trusts

The expected range statistic is used to arrive at a judgement of how a trust is performing for each question that is able to be scored compared with most other trusts that took part in the survey.

If you want to use the scored data in another way, to compare scores between different trusts, you will need to undertake an appropriate statistical test to ensure that any changes are 'statistically significant'. 'Statistically significant' means that you can be very confident that any difference between scores is real and not due to chance.

Although surveys of Emergency Departments have been carried out previously (in [2003, 2004/5, 2008, 2012 and 2014](#)) results from the 2016 survey are **not comparable** with these because of changes made to the sampling strategy. For more detailed information please see the Statistical Release, or the Quality & Methodology report.

6. Further information

The results for England and trust level results can be found on the CQC website. You can also find documents here detailing the trust level and England level analysis methodology:

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 & 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

Appendix A: Scoring for the 2016 Emergency Department survey

The following describes the scoring system applied to the evaluative questions in the survey. Taking question 15 as an example (Figure A1), it asks respondents whether doctors and nurses listened to patients. The option of “No” was allocated a score of 0, as this suggests that the experiences of the patient need to be improved. A score of 10 was assigned to the option ‘Yes, completely’, as it reflects a positive patient experience. The remaining option, ‘Yes, to some extent’, was assigned a score of 5 as the patient felt their fears were somewhat discussed – either sometimes but not always, or some aspects of their anxieties were discussed but not others. Hence it was placed on the midpoint of the scale.

If the patient did not have any anxieties or fears, this was classified as a ‘not applicable’ response, as this option was not a direct measure of the explanations that had been given.

Figure A1 Scoring example:

Question 15 (2016 Emergency Department Survey)

Q15. If you had any anxieties or fears about your condition or treatment, did a doctor or nurse discuss them with you?

Yes, completely	10
Yes, to some extent	5
No	0
I did not have anxieties or fears	Not applicable

Where a number of options lay between the negative and positive responses, they were placed at equal intervals along the scale. For example, question 33 asks respondents how clean the Emergency Department was (Figure A2). The following response options were provided:

- Very clean
- Fairly clean
- Not very clean
- Not at all clean
- Can't say

A score of 10 was assigned to the option ‘Very clean’, as this represents best outcome in terms of patient experience. A response that the Emergency Department was ‘not at all clean’ was given a score of 0. The remaining two answers were assigned a score that reflected their position in terms of quality of experience, spread evenly across the scale. Hence the option ‘fairly clean’ was assigned a score of 6.7, and ‘not very clean’ was given a score of 3.3.

Figure A2 Scoring example:

Question 33 (2016 Emergency Department Survey)

Q33. In your opinion, how clean was the emergency department?

Very clean	10
Fairly clean	6.7
Not very clean	3.3
Not at all clean	0
Can't say	Not applicable

Details of the method used to calculate the scores for each trust, for individual questions and each section of the questionnaire, are available in Appendix B. This also includes an explanation of the technique used to identify scores that are better, worse or about the same as most other trusts.

All analysis is carried out on a 'cleaned' data set. 'Cleaning' refers to the editing process that is undertaken on the survey data. As part of the cleaning process, responses are removed from any trust that has fewer than 30 respondents to a question. This is because the uncertainty around the result is too high, and very low numbers would risk respondents being recognised from their responses. For more information please see the [data cleaning document](#).

The below details the scoring allocated to each scorable question. The sections relate to how questions are ordered in the trust benchmark reports and the data on the CQC website which does not necessarily follow the questionnaire structure: Q31 and Q32 on pain relief are included in the 'care and treatment' section, and Q44 and Q45 are reported in separate sections.

Section 1: Arrival at the emergency department

Q5. Once you arrived at the hospital, how long did you wait with the ambulance crew before your care was handed over to emergency department staff?	
I did not have to wait	10
Up to 15 minutes	10
16 - 30 minutes	6.7
31 - 60 minutes	3.3
More than 1 hour but no more than 2 hours	0
More than 2 hours	0
Don't know/ can't remember	-

Answered by those who arrived by ambulance

Q7: Were you given enough privacy when discussing your condition with the receptionist?	
Yes, definitely	10
Yes, to some extent	5
No	0
I did not discuss my condition with a receptionist	-

Answered by all

Section 2: Waiting times

Q8: How long did you wait before you first spoke to a nurse or doctor?	
0 -15 minutes	10
16-30 minutes	6.7
31-60 minutes	3.3
More than 60 minutes	0
Don't know/ can't remember	-

Answered by all

Q9: Sometimes, people will first talk to a nurse or doctor and be examined later. From the time you arrived, how long did you wait before being examined by a doctor or nurse?	
I did not have to wait	10
1 - 30 minutes	8
31 - 60 minutes	6
More than 1 hour but no more than 2 hours	4
More than 2 hours but no more than 4 hours	2
More than 4 hours	0
Can't remember	-
I did not see a doctor or a nurse	-

Answered by all

Q10: Were you told how long you would have to wait to be examined?	
Yes, but the wait was shorter	10
Yes, and I had to wait about as long as I was told	10
Yes, but the wait was longer	5
No, I was not told	0
Don't know/ can't remember	-

Answered by those who waited to see a doctor or nurse

Q11: Overall, how long did your visit to the Emergency Department last?	
Up to 1 hour	10
More than 1 hour but no more than 2 hours	10
More than 2 hours but no more than 4 hours	8
More than 4 hours but no more than 6 hours	6
More than 6 hours but no more than 8 hours	4
More than 8 hours but no more than 12 hours	2
More than 12 hours but no more than 24 hours	0
More than 24 hours	0
Can't remember	-

Answered by all

Section 3: Doctors and nurses

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	-

Answered by all

Q13: While you were in the emergency department, did a doctor or nurse explain your condition and treatment in a way you could understand?	
Yes, completely	10
Yes, to some extent	5
No	0
I did not need an explanation	-

Answered by those who saw a doctor or nurse

Q14: Did the doctors and nurses listen to what you had to say?	
Yes, definitely	10
Yes, to some extent	5
No	0

Answered by those who saw a doctor or nurse

Q15: If you had any anxieties or fears about your condition or treatment, did a doctor or nurse discuss them with you?	
Yes, completely	10
Yes, to some extent	5
No	0
I did not have anxieties or fears	-

Answered by those who saw a doctor or nurse

Q16: Did you have confidence and trust in the doctors and nurses examining and treating you?	
Yes, definitely	10
Yes, to some extent	5
No	0

Answered by those who saw a doctor or nurse

Q17: Did doctors or nurses talk to each other about you as if you weren't there?	
Yes, definitely	0
Yes, to some extent	5
No	10

Answered by those who saw a doctor or nurse

Q18: If your family or someone else close to you wanted to talk to a doctor, did they have enough opportunity to do so?	
Yes, definitely	10
Yes, to some extent	5
No	0
No family or friends were involved	-
My family or friends did not want or need information	-
I did not want my family or friends to talk to a doctor	-

Answered by those who saw a doctor or nurse

Section 4: Care and treatment

Q19: While you were in the emergency department, how much information about your condition or treatment was given to you?	
Not enough	5
Right amount	10
Too much	5
I was not given any information about my condition or treatment	0

Answered by all

Q20: Were you given enough privacy when being examined or treated?	
Yes, definitely	10
Yes, to some extent	5
No	0

Answered by all

Q21: If you needed attention, were you able to get a member of medical or nursing staff to help you?	
Yes, always	10
Yes, sometimes	5
No, I could not find a member of staff to help me	0
A member of staff was with me all the time	10
I did not need attention	-

Answered by all

Q22: Sometimes in a hospital, a member of staff will say one thing and another will say something quite different. Did this happen to you in the emergency department?	
Yes, definitely	0
Yes, to some extent	5
No	10

Answered by all

Q23: Were you involved as much as you wanted to be in decisions about your care and treatment?	
Yes, definitely	10
Yes, to some extent	5
No	0
I was not well enough to be involved in decisions about my care	-

Answered by all

Q24: If you were feeling distressed while you were in the emergency department, did a member of staff help to reassure you?	
Yes, definitely	10
Yes, to some extent	5
No	0
I was not distressed	-
Not sure/ can't remember	-

Answered by all

Q31: How many minutes after you requested pain relief medication did it take before you got it?	
0 minutes / right away	10
1 - 5 minutes	10
6 - 10 minutes	7.5
11 - 15 minutes	5
16 - 30 minutes	2.5
More than 30 minutes	0
I asked for pain relief medication but wasn't given any	0

Answered by those who were in pain and requested pain relief

Q32: Do you think the hospital staff did everything they could to help control your pain?	
Yes, definitely	10
Yes, to some extent	5
No	0
Can't say/ don't know	-

Answered by those who were in pain

Section 5: Tests

Q26: Did a member of staff explain why you needed these test(s) in a way you could understand?	
Yes, completely	10
Yes, to some extent	5
No	0

Answered by those who had tests

Q28: Did a member of staff explain the results of the tests in a way you could understand?	
Yes, definitely	10
Yes, to some extent	5
No	0
Not sure/ can't remember	-

Answered by those who received their test results before they left the emergency department

Section 6: Hospital environment and facilities

Q33: In your opinion, how clean was the emergency department?	
Very clean	10
Fairly clean	6.7
Not very clean	3.3
Not at all clean	0
Can't say	-

Answered by all

Q34: While you were in the emergency department, did you feel threatened by other patients or visitors?	
Yes, definitely	0
Yes, to some extent	5
No	10

Answered by all

Q35: Were you able to get suitable food or drinks when you were in the emergency department?	
Yes	10
No	0
I was told not to eat or drink	10
I did not know if I was allowed to eat or drink	0
I did not want anything to eat or drink	-

Answered by all

Section 7: Leaving the emergency department

Q38: Did a member of staff explain the purpose of the medications you were to take at home in a way you could understand?	
Yes, completely	10
Yes, to some extent	5
No	0
I did not need an explanation	-

Answered by those who were not admitted to hospital or a nursing home and who were prescribed medication

Q39: Did a member of staff tell you about medication side effects to watch for?	
Yes, completely	10
Yes, to some extent	5
No	0
I did not need this type of information	-

Answered by those who were not admitted to hospital or a nursing home and who were prescribed medication

Q40: Did a member of staff tell you when you could resume your usual activities, such as when to go back to work or drive a car?	
Yes, definitely	10
Yes, to some extent	5
No	0
I did not need this type of information	-

Answered by those who were not admitted to hospital or a nursing home

Q41. Did hospital staff take your family or home situation into account when you were leaving the emergency department?	
Yes, completely	10
Yes, to some extent	5
No	0
It was not necessary	-
Don't know / can't remember	-

Q42: Did a member of staff tell you about what danger signals regarding your illness or treatment to watch for after you went home?	
Yes, completely	10
Yes, to some extent	5
No	0
I did not need this type of information	-

Answered by those who were not admitted to hospital or a nursing home

Q43: Did hospital staff tell you who to contact if you were worried about your condition or treatment after you left the emergency department?	
Yes	10
No	0
Don't know/ can't remember	-

Answered by those who were not admitted to hospital or a nursing home

Section 8: Respect and dignity

Q44: Overall, did you feel you were treated with respect and dignity while you were in the emergency department?	
Yes, all of the time	10
Yes, some of the time	5
No	0

Answered by all

Section 9: Experience overall

Q45: Overall...	
0 (I had a very poor experience)	0
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10 (I had a very good experience)	10

Answered by all

Appendix B: Calculating the trust score and category

Calculating trust scores

The scores for each question and section in each trust were calculated using the method described below.

Weights were calculated to adjust for any variation between trusts that resulted from differences in the age and sex of respondents. A weight was calculated for each respondent by dividing the national proportion of respondents in their age/sex group by the corresponding trust proportion. The reason for weighting the data was that younger people and women tend to be more critical in their responses than older people and men. If a trust had a large population of young people or women, their performance might be judged more harshly than if there was a more consistent distribution of respondents by age and sex.

Please note that as detailed in the introduction (see section 1), this technical document relates to **Type 1 department data** only.

Weighting survey responses

The first stage of the analysis involved calculating national age/ sex proportions. It must be noted that the term “national proportion” is used loosely here as it was obtained from pooling the survey data from all trusts, and was therefore based on the respondent population rather than the entire population of England.

The questionnaire asked respondents to state their year of birth. The approximate age of each patient was then calculated by subtracting the figure given from 2016. The respondents were then grouped according to the categories shown in Figure B1.

If a patient did not fill in their year of birth or sex on the questionnaire, this information was inputted from the sample file. If information on a respondent’s age and/or sex was missing from both the questionnaire and the sample file, the patient was excluded from the analysis.

The national age/sex proportions relate to the proportion of men and women of different age groups. As shown in Figure B1, the proportion of respondents who were male, and aged 51 to 65 years is 0.115; the proportion who were women and aged 51 to 65 years is 0.135, etc.

Figure B1 National Proportions:

Sex	Age Group	National proportion 2016 (Type 1)
Men	≤35	0.045
	36-50	0.061
	51-65	0.115
	66+	0.227
Women	≤35	0.076
	36-50	0.078
	51-65	0.135
	66+	0.263

These proportions were calculated for each trust, using the same procedure.

The next step was to calculate the weighting for each individual. Age/sex weightings were calculated for each respondent by dividing the national proportion of respondents in their age/sex group by the corresponding trust proportion.

If, for example, a lower proportion of men aged between 51 and 65 years within Trust A responded to the survey, in comparison with the national proportion, then this group would be under-represented in the final scores. Dividing the national proportion by the trust proportion results in a weighting greater than “1” for members of this group (Figure B2). This increases the influence of responses made by respondents within that group in the final score, thus counteracting the low representation.

Figure B2 Proportion and Weighting for Trust A

Sex	Age Group	National proportion 2016 (Type 1)	Trust A Proportion	Trust A Weight (National/Trust A)
Men	≤35	0.045	0.036	1.250
	36-50	0.061	0.071	0.859
	51-65	0.115	0.094	1.223
	66+	0.227	0.189	1.201
Women	≤35	0.076	0.092	0.826
	36-50	0.078	0.114	0.684
	51-65	0.135	0.168	0.804
	66+	0.263	0.236	1.114

Likewise, if a considerably higher proportion of women aged between 36 and 50 years from Trust B responded to the survey (Figure B3), then this group would be over-represented within the sample, compared with national representation of this group. Subsequently this group would have a greater influence over the final score. To counteract this, dividing the national proportion by the proportion for Trust B, results in a weighting of less than one for this group.

Figure B3 Proportion and Weighting for Trust B

Sex	Age Group	National proportion 2016 (Type 1)	Trust B Proportion	Trust B Weight (National/Trust B)
Men	≤35	0.045	0.032	1.406
	36-50	0.061	0.058	1.052
	51-65	0.115	0.124	0.927
	66+	0.227	0.188	1.207
Women	≤35	0.076	0.068	1.118
	36-50	0.078	0.207	0.377
	51-65	0.135	0.112	1.205
	66+	0.263	0.211	1.246

To prevent the possibility of excessive weight being given to respondents in an extremely underrepresented group, the maximum value for any weight was set at five.

Calculating question scores

The trust score for each question displayed on the website was calculated by applying the weighting for each respondent to the scores allocated to each response.

The responses given by each respondent were entered into a dataset using the 0-10 scale described in section 3. Each row corresponded to an individual respondent, and each column related to a survey question. For those questions that the respondent did not answer (or received a “not applicable” score for), the relevant cell remained empty. Alongside these were the weightings allocated to each respondent (Figure B4).

Figure B4: Scoring for ‘Hospital environment and facilities’ Trust B

Respondent	Scores			Weight
	Q33	Q34	Q35	
1	10	.	0	1.406
2	6.7	10	10	0.377
3	3.3	5	0	1.246

Respondents’ scores for each question were then multiplied individually by the relevant weighting, in order to obtain the numerators for the trust scores (Figure B5).

Figure B5: Numerators for ‘Hospital environment and facilities’ Trust B

Respondent	Scores			Weight
	Q33	Q34	Q35	
1	14.060	.	0.000	1.406
2	2.526	3.770	3.770	0.377
3	4.112	6.230	0.000	1.246

Obtaining the denominators for each domain score

A second dataset was then created. This contained a column for each question, grouped into domains, and again with each row corresponding to an individual respondent. A value of one was entered for the questions where a response had been given by the respondent, and all questions that had been left unanswered or allocated a scoring of “not applicable” were set to missing (Figure B6).

Figure B6 Values for non-missing responses ‘Hospital environment and facilities’ Trust B

Respondent	Scores			Weight
	Q33	Q34	Q35	
1	1	.	1	1.406
2	1	1	1	0.377
3	1	1	1	1.246

The denominators were calculated by multiplying each of the cells within the second dataset by the weighting allocated to each respondent. This resulted in a figure for each question that the respondent had answered (Figure B7). Again, the cells relating to the questions that the respondent did not answer (or received a ‘not applicable’ score for) remained set to missing.

Figure B7 Denominators for ‘Hospital environment and facilities’ Trust B

Respondent	Score			Weight
	Q33	Q34	Q35	
1	1.406	.	1.406	1.406
2	0.377	0.377	0.377	0.377
3	1.246	1.246	1.246	1.246

The weighted mean score for each trust, for each question, was calculated by dividing the sum of the weighted scores for a question (i.e. numerators), by the weighted sum of all eligible respondents to the question (i.e. denominators) for each trust.

Using the example data for Trust B, we first calculated weighted mean scores for the two questions that contributed to the ‘arrival at A&E’ section of the questionnaire.

$$\text{Q33: } \frac{14.06 + 2.526 + 4.112}{1.406 + 0.377 + 1.246} = 6.833$$

$$\text{Q43: } \frac{3.770 + 6.230}{0.377 + 1.246} = 6.161$$

$$\text{Q35: } \frac{0.000 + 3.770 + 0.000}{1.406 + 0.377 + 1.246} = 1.245$$

Calculating section scores

A simple arithmetic mean of each trust’s question scores was then taken to give the score for each section. Continuing the example from above, then, Trust B’s score for the ‘Arrival at A&E’ section of the A&E survey would be calculated as:

$$(6.833 + 6.161 + 1.245) / 3 = 4.746$$

Calculation of the expected ranges

Z statistics (or Z scores) are standardized scores derived from normally distributed data, where the value of the Z score translates directly to a p-value. That p-value then translates to what level of confidence you have in saying that a value is significantly different from the mean of your data (or your ‘target’ value).

A standard Z score for a given item is calculated as:

$$z_i = \frac{y_i - \theta_0}{s_i} \quad (1)$$

where: s_i is the standard error of the trust score⁴,
 y_i is the trust score
 θ_0 is the mean score for all trusts

Under this banding scheme, a trust with a Z score of < -1.96 is labeled as “Worse” (significantly below average; $p < 0.025$ that the trust score is below the national average), $-1.96 < Z < 1.96$ as “About the same”, and $Z > 1.96$ as “Better”

⁴ Calculated using the method in Appendix C.

(significantly above average; $p < 0.025$ that the trust score is above the national average) than what would be expected based on the national distribution of trust scores.

However, for measures where there is a high level of precision (the survey indicators sample sizes average around 400 to 500 per trust) in the estimates, the standard Z score may give a disproportionately high number of trusts in the significantly above/below average bands (because s_i is generally so small). This is compounded by the fact that all the factors that may affect a trust's score cannot be controlled. For example, if trust scores are closely related to economic deprivation then there may be significant variation between trusts due to this factor, not necessarily due to factors within the trusts' control. In this situation, the data are said to be 'over dispersed'. That problem can be partially overcome by the use of an 'additive random effects model' to calculate the Z score (we refer to this modified Z score as the Z_D score). Under that model, we accept that there is natural variation between trust scores, and this variation is then taken into account by adding this to the trust's local standard error in the denominator of (1). In effect, rather than comparing each trust simply to one national target value, we are comparing them to a national distribution.

The Z_D score for each question and section was calculated as the trust score minus the national mean score, divided by the standard error of the trust score plus the variance of the scores between trusts. This method of calculating a Z_D score differs from the standard method of calculating a Z score in that it recognizes that there is likely to be natural variation between trusts which one should expect, and accept. Rather than comparing each trust to one point only (i.e. the national mean score), it compares each trust to a distribution of acceptable scores. This is achieved by adding some of the variance of the scores between trusts to the denominator.

The steps taken to calculate Z_D scores, based on the method presented in Spiegelhalter et al. (2012), are outlined below.

Winsorising Z-scores

The first step when calculating Z_D is to 'Winsorise' the standard Z scores (from (1)). Winsorising consists of shrinking in the extreme Z-scores to some selected percentile, using the following method:

1. Rank cases according to their naive Z-scores.
2. Identify Z_q and $Z_{(1-q)}$, the 100q% most extreme top and bottom naive Z-scores. For this work, we used a value of $q=0.1$.
3. Set the lowest 100q% of Z-scores to Z_q , and the highest 100q% of Z-scores to $Z_{(1-q)}$. These are the Winsorised statistics.

This retains the same number of Z-scores but discounts the influence of outliers.

Estimation of over-dispersion

An over dispersion factor $\hat{\phi}$ is estimated for each indicator which allows us to say if the data for that indicator are over dispersed or not:

$$\hat{\phi} = \frac{1}{I} \sum_{i=1}^I z_i^2 \quad (2)$$

where I is the sample size (number of trusts) and z_i is the Z score for the i th trust given by (1). The Winsorised Z scores are used in estimating $\hat{\phi}$.

An additive random effects model

If $I \hat{\phi}$ is greater than $(I - 1)$ then we need to estimate the expected variance between trusts. We take this as the standard deviation of the distribution of θ_i (trust means) for trusts, which are on target, we give this value the symbol $\hat{\tau}$, which is estimated using the following formula:

$$\hat{\tau}^2 = \frac{I\hat{\phi} - (I - 1)}{\sum_i w_i - \sum_i w_i^2 / \sum_i w_i} \quad (3)$$

where $w_i = 1 / s_i^2$ and $\hat{\phi}$ is from (2). Once $\hat{\tau}$ has been estimated, the Z_D score is calculated as:

$$Z_i^D = \frac{y_i - \theta_0}{\sqrt{s_i^2 + \hat{\tau}^2}} \quad (4)$$

References

Spiegelhalter D., Sherlaw-Johnson, C., Bardsley, M., Blunt, I., Wood, C., & Grigg, O. (2012). Statistical methods for healthcare regulation: Rating, screening and surveillance. *Journal of the Royal Statistical Society (Series A)*, 175(1), 1-47.

Appendix C: Calculation of standard errors

In order to calculate statistical bandings from the data, it is necessary for CQC to have both trusts' scores for each question and section and the associated standard error. Since each section is based on an aggregation of question mean scores that are based on question responses, a standard error needs to be calculated using an appropriate methodology.

For the patient experience surveys, the z-scores are scores calculated for section and question scores, which combines relevant questions making up each section into one overall score, and uses the pooled variance of the question scores

Assumptions and notation

The following notation will be used in formulae:

X_{ijk}	is the score for respondent j in trust i to question k
Q	is the number of questions within section d
w_{ij}	is the standardization weight calculated for respondent j in trust i
Y_{ik}	is the overall trust i score for question k
Y_{id}	is the overall score for section d for trust i

Associated with the subject or respondent is a weight w_{ij} corresponding to how well the respondent's age/sex is represented in the survey compared with the population of interest.

Calculating mean scores

Given the notation described above, it follows that the overall score for trust i on question k is given as:

$$Y_{ik} = \frac{\sum_j w_{ij} X_{ijk}}{\sum_j w_{ij}}$$

The overall score for section d for trust i is then the average of the trust-level question means within section d . This is given as:

$$Y_{id} = \frac{\sum_{k=1}^Q Y_{ikd}}{Q}$$

Calculating standard errors

Standard errors are calculated for both sections and questions.

The variance within trust i on question k is given by:

$$\hat{\sigma}_{ik}^2 = \frac{\sum_j w_{ij} \left(X_{ijk} - Y_{ik} \right)^2}{\sum_j w_{ij}}$$

This assumes independence between respondents.

For ease of calculation, and as the sample size is large, we have used the biased estimate for variance.

The variance of the trust level average question score, is then given by:

$$\begin{aligned} V_{ik} &= \text{Var}(Y_{ik}) = \text{Var} \left(\frac{\sum_j w_{ij} X_{ijk}}{\sum_j w_{ij}} \right) \\ &= \frac{\text{Var} \left(\sum_j w_{ij} X_{ijk} \right)}{\left(\sum_j w_{ij} \right)^2} \\ &= \frac{\hat{\sigma}_{ik}^2 \sum_j w_{ij}^2}{\left(\sum_j w_{ij} \right)^2} \end{aligned}$$

Covariances between pairs of questions (here, k and m) can be calculated in a similar way:

$$COV_{ik.im} = \text{Cov}(Y_{ik}, Y_{im}) = \frac{\hat{\sigma}_{ikm} \sum_j w_{ij}^2}{\left(\sum_j w_{ij} \right)^2}$$

$$\text{Where } \hat{\sigma}_{ikm} = \frac{\sum_j w_{ij} (X_{ijk} - Y_{ik})(X_{ijm} - Y_{im})}{\sum_j w_{ij}}$$

Note: w_{ij} is set to zero in cases where patient j in trust i did not answer both questions k and m .

The trust level variance for the section score d for trust i is given by:

$$V_{id} = \text{Var}(Y_{id}) = \frac{1}{Q^2} \left\{ \sum_{k=1}^Q V_{ik} + 2 \sum_{k=2}^Q \sum_{m=1}^{k-1} \text{COV}_{ik,im} \right\}$$

The standard error of the section score is then:

$$SE_{id} = \sqrt{V_{id}}$$