Care Quality Commission (CQC)

Technical details – patient survey information
2012 Accident and Emergency Survey
December 2012

Contents

1. Introduction

This document outlines the methods used by the Care Quality Commission to score and analyse the results for the 2012 Accident and Emergency (A&E) Survey, as available on the Care Quality Commission website, and in the benchmark report for each trust.

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. An A-to-Z list of trust names is available at the link below, containing further links to the survey data for all NHS trusts that took part in the survey:
www.cqc.org.uk/accidentandemergency

The CQC webpage also contains the national results for England, comparing against the results for the previous survey. A link to the benchmark report for each trust is also here, linking though to where these are held on the patient survey coordination centre website. Results displayed in the benchmark report for each trust are a graphical representation of the results displayed for the public on the CQC website (see further information section). These provide more detailed information for NHS acute trusts who took part in the survey.

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 'Scoring individual questions' for more detail). Questions that are not presented in this way tend to be those included solely for 'filtering' respondents past any questions that may not be relevant to them (such as: 'Were you taken to the hospital in an ambulance?') or those used for descriptive or information purposes.

The scores for each question are generally grouped on the website according to the sections of the questionnaire as completed by respondents. For example, the A&E survey includes sections on 'doctors and nurses,' 'care and treatment' and 'tests', amongst others. The average score for each trust, for each section, is also calculated and presented on the website.

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

- Better (the trust is performing 'better' compared with most other trusts in the survey)
- About the same (the trust is performing 'about the same' as most other trusts in the survey)
- Worse (the trust is performing 'worse' compared with most other trusts in the survey)

3. The CQC organisation search tool

The organisation search tool was previously referred to as the Care Directory, and survey data has been displayed in it since 2007. It is intended for a public audience, and contains information from various areas within the Care Quality Commission's

functions. 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 most other trusts.

The survey data can be found from the A to Z link available at:
www.cqc.org.uk/accidentandemergency

Or by searching for a trust from the CQC home page, then clicking on 'Patient survey information' on the right hand side, or searching for an NHS trust, then selecting the survey under the 'Reports and surveys about this organisation' tab.

4. Interpreting the data

4.1 Scoring

The 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 patient'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 possible reported patient experience. 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 trusts' performance in terms of A&E patient 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 assigned.

4.2 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 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.

The 2012 A&E Survey is standardised by age and gender.

4.3 Expected range

The better / about the same / worse categories shown on the website 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 B for more details). The red, green and orange sections in the benchmark report charts 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.

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.

4.4 Comparing scores across trusts or across survey years

The expected range statistic is used to arrive at a judgement of a how a trust is performing compared with all other trusts that took part in the survey. However, if you want to use the scored data in another way, to compare scores (either as trend data for an individual trust or 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 and change between scores is real and not due to chance. The benchmark report for each trust includes a comparison to the 2008 survey scores and indicates whether the change is statistically significant. However, to compare back to the 2004 survey (where possible) you would need to undertake a similar significance test.

4.5 Conclusions made on performance

It should be noted that the data only show performance relative to other trusts: there are no absolute thresholds for ‘good’ or ‘bad’ performance. Thus, a trust may score low 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.

The better / worse categories are intended to help trusts identify areas of good or poor performance. However, 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 accurate to look at actual changes in scores and to test for statistically significant differences.

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. The number of questions on each topic in the survey varies, and often so does trusts performance across these. So if you counted across all of them, some topics will have more influence on the overall average than others, when in fact some might not be so important.

5. Further information

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 this technical document):
www.cqc.org.uk/accidentandemergency

The results for previous 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/surveys/626

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

Appendix A: Scoring for the 2012 Accident and Emergency Survey results

The following describes the scoring system applied to the evaluative questions in the survey. Taking question 14 as an example (Figure A1), it asks respondents whether they were able to discuss any anxieties or fears with hospital staff. The option of "No" was allocated a score of 0, as this suggests that the experience of the patient could have been 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 on the midpoint of the scale, as experience was not entirely negative but there was room for some improvement.

If the patient said they 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 communications between staff and patients.

Figure A1 Scoring example:
Question 14 (2012 Accident and Emergency Survey)

Q14. 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 30 asks how clean the toilets in the A&E department were. (Figure A2). The following response options were provided:

- Very clean
- Fairly clean
- Not very clean
- Not at all clean

A score of 10 was assigned to the option 'Very clean', as this represents the best outcome in terms of patient experience. A response of '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 and shown in Figure A2 below.

Figure A2 Scoring example:
Question 30 (2012 Accident and Emergency Survey)

Q30. How clean were the toilets in the A&E Department?

Very clean	10
Fairly clean	6.7
Not very clean	3.3
Not at all clean	0

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.

The below sets out the scoring assigned to each question used in the analysis.

Section 1: Travelling by ambulance

4. How well do you think the ambulance service and A&E staff worked together?

Very well	10
Fairly well	6.7
Not very well	3.3
Not at all well	0
Don't know/ can't remember	Not applicable

Answered by all who travelled to A&E in an ambulance

5. Once you arrived at hospital, how long did you wait with the ambulance crew before your care was handed over to the A&E 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	Not applicable

Answered by all who travelled to A&E in an ambulance

Section 2: Reception and waiting

6. 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	Not applicable

Answered by all

7. 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	Not applicable

Answered by all

8. From the time you first arrived at the A&E Department, 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	Not applicable
I did not see a doctor or a nurse	Not applicable

Answered by all

9. 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	Not applicable

Answered by all who waited to be examined by a doctor or a nurse

10. Overall, how long did your visit to the A&E Department last?

Up to 1 hour	10
More than 1 hour but no more than 2 hours	8.3
More than 2 hours but no more than 4 hours	6.7
More than 4 hours but no more than 6 hours	5
More than 6 hours but no more than 8 hours	5
More than 8 hours but no more than 12 hours	3.3
More than 12 hours but no more than 24 hours	1.7
More than 24 hours	0
Can't remember	Not applicable

Answered by all

Section 3: Doctors and nurses

11. 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

Answered by all

12. While you were in the A&E 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	Not applicable

Answered by all

13. Did the doctors and nurses listen to what you had to say?

Yes, definitely	10
Yes, to some extent	5
No	0

Answered by all

14. 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 any anxieties or fears	Not applicable

Answered by all

15. 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 all

16. Did doctors or nurses talk in front of you as if you weren't there?	
Yes, definitely	0
Yes, to some extent	5
No	10
Answered by all	

17. 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	Not applicable
My family or friends did not want or need information	Not applicable
I did not want my family or friends to talk to a doctor	Not applicable
Answered by all	

Section 5 your care and treatment

18. While you were in the A&E 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	

19. Were you given enough privacy when being examined or treated?	
Yes, definitely	10
Yes, to some extent	5
No	0
Answered by all	

20. 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	Not applicable
Answered by all	

21. 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 A&E Department?	
Yes definitely	0
Yes to some extent	5
No	10
Answered by all	

22. 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	Not applicable
Answered by all	

28. 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	Not applicable
Answered by all	

Section 5: Tests

24. 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 all who had tests while at the A&E Department	

25. Before you left the A&E Department, did you get the results of your tests?	
Yes	10
No	0
I was told that the results of the tests would be given to me at a later date	Not applicable
Don't know/ can't remember	Not applicable
Answered by all who had tests while at the A&E Department	

26. 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	Not applicable

Answered by all who were given the results of their tests before they left A&E

Section 6: Hospital environment and facilities

29. In your opinion, how clean was the A&E Department?	
Very clean	10
Fairly clean	6.7
Not very clean	3.3
Not at all clean	0
Can't say	Not applicable

Answered by all

30. How clean were the toilets in the A&E Department?	
Very clean	10
Fairly clean	6.7
Not very clean	3.3
Not at all clean	0
I did not use a toilet	Not applicable

Answered all

31. While you were in the A&E Department, did you feel threatened by other patients or visitors?	
Yes definitely	0
Yes to some extent	5
No	10

Answered by all

32. Were you able to get suitable food or drinks when you were in the A&E 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	Not applicable

Answered by all

Section 7: Leaving the A&E Department

35. 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	Not applicable

Answered by all who were discharged from hospital and prescribed new medications

36. 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	Not applicable

Answered by all who were discharged from hospital and prescribed new medications

37. 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	Not applicable

Answered by all who were discharged from hospital

38. Did hospital staff take your family or home situation into account when you were leaving the A&E Department?	
Yes, completely	10
Yes, to some extent	5
No	0
It was not necessary	Not applicable
Don't know/ can't remember	Not applicable

Answered by all who were discharged from hospital

39. 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	Not applicable

Answered by all who were discharged from hospital

40. Did hospital staff tell you who to contact if you were worried about your condition or treatment after you left the A&E Department?

Yes	10
No	5
Don't know/ can't remember	Not applicable

Answered by all who were discharged from hospital

41. As far as you know, was your GP given all the necessary information about the treatment or advice that you received in the A&E Department?

Yes	10
No	0
Don't know	Not applicable
No information was needed	Not applicable

Answered by all who were discharged from hospital

Section 8: Overall

42. Overall, did you feel you were treated with respect and dignity while you were in the A&E Department?

Yes, all of the time	10
Yes, some of the time	5
No	0

Answered by all

Q43 Overall.....

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
I had a very good experience	10

Answered by all

44. While in the A&E Department, did you ever see any posters or leaflets explaining how to complain about the care you received?

Yes	10
No	0
Don't know/ can't remember	Not applicable

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 groupings 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 age and sex of respondents.

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 2012. 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 as it is not possible to assign a weight.

The national age/sex proportions relate to the proportion of men, and women of different age groups. As shown in Figure B1 below, the proportion of respondents who were male and aged 51 to 65 years is 0.114; the proportion who were women and aged 51 to 65 years is 0.125, etc.

Figure B1 National Proportions

Sex	Age Group	National proportion 2012
Men	≤35	0.068
	36-50	0.082
	51-65	0.114
	66+	0.183
Women	≤35	0.105
	36-50	0.103
	51-65	0.125
	66+	0.221

Note: All proportions are given to three decimals places for this example. The analysis included these figures to nine decimal places, and can be provided on request from the CQC surveys team at patient.survey@cqc.org.uk.

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 who were 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 one 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	Trust A Proportion	Trust A Weight (National/Trust A)
Men	≤35	0.068	0.036	1.889
	36-50	0.082	0.071	1.155
	51-65	0.114	0.094	1.213
	66+	0.183	0.189	0.968
Women	≤35	0.105	0.092	1.141
	36-50	0.103	0.114	0.904
	51-65	0.125	0.168	0.744
	66+	0.221	0.236	0.936

Note: All proportions are given to three decimals places for this example. The analysis included these figures to nine decimal places

Likewise, if a considerably higher proportion of women who aged between 36 and 50 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	Trust B Proportion	Trust B Weight (National/Trust B)
Men	≤35	0.068	0.032	2.125
	36-50	0.082	0.058	1.414
	51-65	0.114	0.124	0.919
	66+	0.183	0.188	0.973
Women	≤35	0.105	0.068	1.544
	36-50	0.103	0.207	0.498
	51-65	0.125	0.112	1.116
	66+	0.221	0.211	1.047

Note: All proportions are given to three decimals places for this example. The analysis included these figures to nine decimal places

To prevent the possibility of excessive weight being given to respondents in an extremely under-represented group, the maximum value for any weight was set at five. There was no minimum weight for respondents as applying very small weights to over-represented groups does not have the same potential to give excessive impact to the responses of small numbers of individual respondents.

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 4.1 and outlined in Appendix A. 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 the 'Hospital environment and facilities' section, 2012 Accident and Emergency survey, Trust B

Respondent	Q29	Q30	Q31	Q32	Weight
1	.	6.7	5	10	0.973
2	6.7	10	10	10	1.116
3	.	3.3	0	0	1.047

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 the 'Hospital environment and facilities' section, 2012 Accident and Emergency survey, Trust B

Respondent	Scores				
	Q29	Q30	Q31	Q32	Weight
1	.	6.519	4.865	9.730	0.973
2	7.477	11.160	11.160	11.160	1.116
3	.	3.455	0	0	1.047

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' section, 2012 Accident and Emergency survey, Trust B

Respondent	Scores				
	Q29	Q30	Q31	Q32	Weight
1	.	1	1	1	0.973
2	1	1	1	1	1.116
3	.	1	1	1	1.047

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 the 'Hospital environment and facilities' section, 2012 Accident and Emergency survey, Trust B

Respondent	Scores				
	Q29	Q30	Q31	Q32	Weight
1	.	0.973	0.973	0.973	0.973
2	1.116	1.116	1.116	1.116	1.116
3	.	1.047	1.047	1.047	1.047

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 each of the questions that contributed to the 'hospital environment and facilities' section of the questionnaire.

Q29: $\frac{7.477}{1.116} = 6.7$

Q30: $\frac{6.519+ 11.160+ 3.455}{0.973+ 1.116+ 1.047} = 6.74$

Q31: $\frac{4.865 + 11.160 + 0}{0.973 + 1.116 + 1.047} = 5.11$

Q32: $\frac{9.730 + 11.160 + 0}{0.973 + 1.116 + 1.047} = 6.66$

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 'health and social care section' section of the Community Mental Health Survey would be calculated as:

(6.7 + 6.74 + 5.11 + 6.66) / 4 = 6.303

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).

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.

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

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 score[1],
y_i is the trust score
θ_0 is the mean score for all trusts

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 steps taken to calculate Z_D scores 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=0.2

3. Set the lowest 100q% of Z-scores to Z_q, and the highest 100q% of Z-scores to $_{(1-q)}$. 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 which allows us to say if the data are over dispersed or not:

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

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 $\hat{\phi}$.

[1] Calculated using the method in Appendix C.

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:

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

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:

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

Appendix C: Calculation of standard errors

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:

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

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:

$$Y_{id} = \frac{\sum Y_{ikd}}{Q}$$

Calculating standard errors

Standard errors are calculated for both sections and questions.

The variance of question X_{ijk} at the individual level is given by:

$$V_{ijk} = \frac{\sum w_{ij}\left(X_{ijk} - Y_{ik}\right)^2}{\sum w_{ij}}$$

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:

$$V_{ik} = \frac{\sum w_{ij}\left(X_{ijk} - Y_{ik}\right)^2}{\left(\sum w_{ij}\right)^2}$$

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

$$COV_{ik.im.} = \frac{\sum w_{ij}\left(X_{ijk} - Y_{ik}\right)\left(X_{ijm} - Y_{im}\right)}{\left(\sum w_{ij}\right)^2}$$

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

If questions *k* and *m* comprise a two-item section *d*, then the score for section *d* is a weighted sum of the separate question scores, with each question weighted by ½. The trust level variance for the section score *d* for trust *i* is therefore given by:

$$V_{id} = \frac{V_{ik}}{(2)^2} + \frac{V_{im}}{(2)^2} + 2.\frac{COV_{ik.im}}{(2)^2}.$$

The standard error of the section score is then:

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

This simple case can be extended to cover sections of greater length.