**Care Quality Commission (CQC)**

**Technical details – patient survey information
2013/2014 Ambulance survey of 'Hear and Treat' callers
June 2014**

**Contents**

# 1. Introduction

This document outlines the methods used by the Care Quality Commission to score and analyse the results for the 2013/2014 telephone survey of Ambulance 'Hear and Treat' callers, as available on the Care Quality Commission website and in the benchmark report for each trust.

Trust level results are made available in two ways: on the CQC website, and in a benchmark report for each trust.

On publication of the survey on 8 July 2014, an A-to-Z list of trust names will be available on the CQC website at the link below, containing further links to the survey data for all ambulance trusts that took part in the survey:
**www.cqc.org.uk/Ambulancesurvey201314**

The benchmark reports have been provided to all trusts in advance of the publication, and will be available on the Survey Co-ordination Centre website from 8 July 2014, at: **www.nhssurveys.org**.

The CQC website displays the data in a more simplified way than the benchmark reports, identifying whether a trust performed 'Above', 'Below' or 'Average' when looking across other trusts results for each question and section. This corresponds to the 'Better', 'About the same', and 'Worse' labels used for the same data in the benchmark reports. For more detail on the methodology, please see section five.

The CQC webpage also contains the national results for England as a whole, in the form of the percentage of respondents to each survey question.

# 2. Selecting data for the reporting

Scores are assigned to responses to all survey 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 or those used for descriptive or information purposes.

The scores for each question are grouped on the website and in the benchmark reports according to the sections of the questionnaire that was used by the telephone interviewers. For example, the ambulance survey includes sections on 'the call handler', 'the clinical advisor' amongst others. The average score for each trust, for each section, was calculated and will be presented on the website and in the benchmark report for each trust.

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

- Above
- Average
- Below

(Please note that in the benchmark reports this is labelled as 'better' 'about the same' and '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 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 we'd expect, when compared with most other trusts.

The survey data can be found from the A to Z link available at:
**www.cqc.org.uk/Ambulancesurvey201314**

Or by searching for a trust from the CQC home page, clicking on the NHS trust name, then selecting the survey under the 'Surveys' tab.

## 4. The trust benchmark reports

Benchmark reports should be used by NHS trusts to identify how they are performing in relation to all other trusts that took part in the survey.  From this, areas for improvement can be identified. The reports are available from the Survey Co-ordination Centre website: **www.nhssurveys.org**.

The graphs included in the reports display the scores for a trust, compared with the full range of results from all other trusts that took part in the survey. Each bar represents the range of results for each question across all trusts that took part 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 trust 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.

(Please note that on the CQC website this is labelled as 'above' 'average' and 'below'.)

A black diamond represents the score for a 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

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 callers 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 caller 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.

### 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 callers 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. 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 2014 ambulance survey of 'Hear and Treat' callers is standardised by **age and gender**.

### 5.3 Expected range

The above / average / below categories (better / about the same / worse in the benchmark reports) are based on a statistic called the 'expected range' which determines the range within which the trust's score could fall without differing significantly from the average, taking into account the number of respondents for each trust and the scores for all other trusts. If the trust's performance is outside of this range, it means that it performs significantly above/below what would be expected. If it is within this range, we say that its performance is 'about the same'. This means that where a trust is performing 'better' or 'worse' than the majority of other trusts, it is very unlikely to have occurred by chance.

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 Comparing scores across or within trusts

The expected range statistic is used to arrive at a judgement of 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 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.

**5.5 Conclusions made on performance**

It should be noted that the data only shows performance relative to other trusts: there are no 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 a region' *overall*. Adding up the number of 'better' (above) and 'worse' (below) 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. It also does not take into account the relative importance of each question, for example, one trust may do well on several questions asking about minor issues, but another trust may do well on fewer questions asking about more important issues. It is therefore better to consider which questions or topics may be of interest to you.

**6. Further information**

The full national results are on the CQC website, together with an A to Z list to view the results for each ambulance trust (alongside the technical document outlining the methodology and the scoring applied to each question):
www.cqc.org.uk/Ambulancesurvey201314

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

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

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

**Appendix A: Scoring for the 2013/2014 Ambulance survey of 'Hear and Treat' callers**

The following describes the scoring system applied to the evaluative questions in the survey. Taking the question called ListenCT as an example (Figure A1), it asks respondents whether the first person they spoke to listened to what they had to say. The option of "No" was allocated a score of 0, as this suggests that the experiences of the caller need to be improved. A score of 10 was assigned to the option 'Yes, completely', as it reflects a positive experience. The remaining option, 'Yes, to some extent', was assigned a score of 5 as the caller did not always feel listened to. Hence it was placed on the midpoint of the scale.

If the caller said that they did not know, this was classified as a 'not applicable' response, as this option was not a direct measure of trust staff and was therefore not scored.

**Figure A1 Scoring example:**

**2013/2014 Ambulance survey of 'Hear and Treat' callers**

**ListenCT.**

**Did they listen to what you had to say?**

| | |
|---|---|
| Yes, completely | 10 |
| Yes, to some extent | 5 |
| No | 0 |
| Don't know | Not scored |

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.

The below details the scoring allocated to each scorable question. Please note that due to the extensive filtering that was incorporated into the interview schedule for the telephone survey, question numbering has not been added, as the process was not linear for all respondents. Instead, we have reported results using abbreviated question names, for example: "WhyNotFollow" represents the question "Why was it not possible to follow the advice?"

**Section 1: The Call Handler**

**ListenCT.**

**Did they listen to what you had to say? [first person spoken with]**

| | |
|---|---|
| Yes, completely | 10 |
| Yes, to some extent | 5 |
| No | 0 |
| Don't know | Not scored |
| Answered by all | |

**ReassureCT.**

**Was the [first] person you spoke with reassuring?**

| | |
|---|---|
| Yes, completely | 10 |
| Yes, to some extent | 5 |

| No | 0 |
|---|---|
| Not sure / can't remember | Not scored |
| Answered by all | |

**ConfCT.**
**Did you have confidence in the [first] person you spoke with?**

| Yes, completely | 10 |
|---|---|
| Yes, to some extent | 5 |
| No | 0 |
| Don't know | Not scored |
| Answered by all | |

**DignityCT.**
**Were you treated with dignity and respect by the [first] person you spoke with?**

| Yes, completely | 10 |
|---|---|
| Yes, to some extent | 5 |
| No | 0 |
| Don't know | Not scored |
| Answered by all | |

**UnderstandingCT.**
**Did you feel that the [first] person you spoke to understood what you were telling them?**

| Yes, completely | 10 |
|---|---|
| Yes, to some extent | 5 |
| No | 0 |
| Not sure / can't remember | Not scored |
| Answered by all | |

**InfoGivenCT.**
**Did the first person tell you what to do if the situation changed?**

| Yes | 10 |
|---|---|
| No | 0 |
| Not sure / can't remember | Not scored |
| Answered by all who spoke to two or more people at the ambulance service | |

**InfoFollowCT.**
**Were you able to understand the instructions given?**

| Yes, Completely | 10 |
|---|---|
| Yes, to some extent | 5 |
| No | 0 |
| Not given any instructions | Not scored |
| Not sure / can't remember | Not scored |
| Answered by all who spoke to two or more people at the ambulance service **and** were told what to do if the situation changed | |

**AwareCallBackTA.**
**Were you told you would receive a call back?**

| Yes, you were told | 10 |
|---|---|
| No, you were not told and would have liked to have known | 0 |
| No, you were not told and you did not mind | 0 |
| Not sure / can't remember | Not scored |

Answered by all who spoke with a second person at the ambulance service **and** were called back after the original call

**WaitCallBackTA.**
**Were you told what to do while you waited for the call back?**

| | |
|---|---|
| Yes, you were told | 10 |
| No, you were not told and would have liked to have known | 0 |
| No, you were not told and you did not mind | 0 |
| Not sure / can't remember | Not scored |

Answered by all who spoke with a second person at the ambulance service **and** were told they would be called back

**ExplainTimeCallBack.**
**Were you told when you would be called back?**

| | |
|---|---|
| Yes, you was told | 10 |
| No, you were not told and would have liked to have known | 0 |
| No, you were not told and you did not mind | 0 |
| Not sure / can't remember | Not scored |

Answered by all who spoke with a second person at the ambulance service **and** were told they would be called back

## Section 2: The Clinical Advisor

**FeelCallBackTA.**
**Which of the following best describes how you feel about the length of time you waited before someone called you back?**

| | |
|---|---|
| It was sooner than I expected | 10 |
| It was as soon as I thought was necessary | 10 |
| It should have been a bit sooner | 5 |
| It should have been a lot sooner | 0 |
| Not sure / can't remember | Not scored |

Answered by all who spoke with a second person at the ambulance service **and** were called back after the original call

**ExplainTA.**
**Did the ambulance service explain why an ambulance would not be sent on this occasion?**

| | |
|---|---|
| Yes, completely | 10 |
| Yes, to some extent | 5 |
| No | 0 |
| Not applicable | Not scored |
| Not sure / can't remember | Not scored |

Answered by all who spoke with a second person at the ambulance service

**ListenTA.**
**Did the second person you spoke to at the ambulance service listen to what you had to say?**

| | |
|---|---|
| Yes, completely | 10 |
| Yes, to some extent | 5 |
| No | 0 |
| Not sure / can't remember | Not scored |

Answered by all who spoke with a second person at the ambulance service

**FearsTA.**
**Did you have the opportunity to discuss any fears or anxieties with the second person you spoke with?**

| | |
|---|---|
| Yes, and I did this | 10 |
| I could have done but I didn't want to | 10 |
| No | 0 |
| Did not have any anxieties or fears | Not Scored |
| Not sure / can't remember | Not scored |

Answered by all who spoke with a second person at the ambulance service

**ReassureTA.**
**Was the second person you spoke with reassuring?**

| | |
|---|---|
| Yes, completely | 10 |
| Yes, to some extent | 5 |
| No | 0 |
| Not sure / can't remember | Not scored |

Answered by all who spoke with a second person at the ambulance service

**ConfTA.**
**Did you have confidence in the second person you spoke with?**

| | |
|---|---|
| Yes, completely | 10 |
| Yes, to some extent | 5 |
| No | 0 |
| Not sure / can't remember | Not scored |

Answered by all who spoke with a second person at the ambulance service

**DignityTA.**
**Were you treated with dignity and respect by the second person you spoke with?**

| | |
|---|---|
| Yes, completely | 10 |
| Yes, to some extent | 5 |
| No | 0 |
| Not sure / can't remember | Not scored |

Answered by all who spoke with a second person at the ambulance service

**Section 3: Outcome**

**UnderstandAdvice.**
**Did you understand the advice given at the end of your call with the ambulance service?**

| | |
|---|---|
| Yes, completely | 10 |
| Yes, to some extent | 5 |
| No | 0 |
| Not sure / can't remember | Not scored |

Answered by callers for whom an ambulance was not sent

**ExplainAdv.**
**Did the ambulance service explain the reason for this advice?**

| | |
|---|---|
| Yes, completely | 10 |
| Yes, to some extent | 5 |
| No | 0 |

| Did not want or need an explanation | Not scored |
|---|---|

Answered by callers for whom an ambulance was not sent **and** who remembered whether they agreed with the advice they were given or not

---

**FollowAdvice.**
**Was it possible to follow the advice given?**

| | |
|---|---|
| Yes, completely | 10 |
| Yes, to some extent | 5 |
| No | 0 |
| The advice was for the patient to follow | Not scored |
| It was possible to follow the advice but I chose not to | Not scored |
| Not sure / can't remember | Not scored |

Answered by callers for whom an ambulance was not sent **and** who remembered whether the ambulance service explained the reason for the advice

---

**ExplainAmb.**
**Did the ambulance service explain why an ambulance would not be sent on this occasion?**

| | |
|---|---|
| Yes, completely | 10 |
| Yes, to some extent | 5 |
| No | 0 |
| Not sure / can't remember | Not scored |

Answered by callers for whom an ambulance was not sent **and** only spoke to one person at the ambulance service

---

## Section 4: Overall

**OverDignity.**
**Overall, did you feel you were treated with respect and dignity by the ambulance service?**

| | |
|---|---|
| Yes, always | 10 |
| Yes, sometimes | 5 |
| No | 0 |

Answered by all who spoke to two or more people at the ambulance service

---

**Overall.**
**If you had questions to ask ambulance staff did you have the opportunity to ask them?**

| | |
|---|---|
| Yes | 10 |
| No | 0 |
| I did not have any questions to ask ambulance staff | Not scored |

Answered by all

---

**OverKind.**
**Overall, were you treated with kindness and understanding by the ambulance service?**

| | |
|---|---|
| Yes, all of the time | 10 |
| Yes, some of the time | 5 |
| No | 0 |
| Don't know | Not scored |

Answered by all

**OverallExp.**
**On a scale of 0 to 10 where 0 is 'I had a very poor experience'**
**and 10 is 'I had a very good experience', how was your overall**
**experience with the ambulance service?**

| | |
|---|---|
| 0 | 0 |
| 1 | 1 |
| 2 | 2 |
| 3 | 3 |
| 4 | 4 |
| 5 | 5 |
| 6 | 6 |
| 7 | 7 |
| 8 | 8 |
| 9 | 9 |
| 10 | 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 is 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 caller was then calculated by subtracting the figure given from 2013. The respondents were then grouped according to the categories shown in Figure B1.

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 0.093; the proportion who were women and aged 51 to 65 years is 0.147, etc.

### Figure B1 National Proportions

| Sex | Age Group | National proportion 2013 |
|---|---|---|
| Men | 18-35 | 0.108 |
| | 36-50 | 0.092 |
| | **51-65** | **0.093** |
| | 66+ | 0.142 |
| Women | 18-35 | 0.120 |
| | 36-50 | 0.109 |
| | **51-65** | **0.147** |
| | 66+ | 0.188 |

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 aged between 36 and 50 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 | Trust A Proportion | Trust A Weight (National/Trust A) |
|-----|-----------|---------------------|--------------------|-----------------------------------|
| Men | ≤35 | 0.108 | 0.036 | 3.000 |
| | 36-50 | 0.092 | 0.071 | 1.296 |
| | 51-65 | 0.093 | 0.094 | 0.989 |
| | 66+ | 0.142 | 0.189 | 0.751 |
| Women | ≤35 | 0.120 | 0.092 | 1.304 |
| | 36-50 | 0.109 | 0.114 | 0.956 |
| | 51-65 | 0.147 | 0.168 | 0.875 |
| | 66+ | 0.188 | 0.236 | 0.797 |

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.

Likewise, if a considerably higher proportion of 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 | Trust B Proportion | Trust B Weight (National/Trust B) |
|-----|-----------|---------------------|--------------------|-----------------------------------|
| Men | ≤35 | 0.108 | 0.032 | 3.375 |
| | 36-50 | 0.092 | 0.058 | 1.586 |
| | 51-65 | 0.093 | 0.124 | 0.750 |
| | 66+ | 0.142 | 0.188 | 0.755 |
| Women | ≤35 | 0.120 | 0.068 | 1.765 |
| | 36-50 | 0.109 | 0.207 | 0.527 |
| | 51-65 | 0.147 | 0.112 | 1.313 |
| | 66+ | 0.188 | 0.211 | 0.891 |

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.

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 the 'Outcome' section, 2013 survey of 'Hear and Treat' callers, Trust B**

| Respondent | Scores | | | | Weight |
| --- | --- | --- | --- | --- | --- |
| | Underst andAdvi ce | Explain Adv | FollowAd vice | ExplainA mb | |
| 1 | 10 | 5 | 5 | 10 | 1.586 |
| 2 | 0 | 10 | 5 | 5 | 0.755 |
| 3 | . | 0 | 10 | . | 1.313 |

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 'Outcome' section, 2013 survey of 'Hear and Treat' callers, Trust B**

| Respondent | Scores | | | | Weight |
| --- | --- | --- | --- | --- | --- |
| | Underst andAdvi ce | Explain Adv | FollowAd vice | ExplainA mb | |
| 1 | 15.86 | 7.93 | 7.93 | 15.86 | 1.586 |
| 2 | 0 | 7.55 | 3.775 | 3.775 | 0.755 |
| 3 | . | 0 | 13.13 | . | 1.313 |

**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, 'Outcome' section, 2013 survey of 'Hear and Treat' callers, Trust B**

| Respondent | Scores | | | | Weight |
| --- | --- | --- | --- | --- | --- |
| | Underst andAdvi ce | Explain Adv | FollowAd vice | ExplainA mb | |
| 1 | 1 | 1 | 1 | 1 | 1.586 |
| 2 | 1 | 1 | 1 | 1 | 0.755 |

| 3 | . | 1 | 1 | . | 1.313 |
|---|---|---|---|---|-------|

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. Again, the cells relating to the questions that the respondent did not answer (or received a 'not applicable' for) remained set to missing (Figure B7).

**Figure B7 Denominators for the 'Outcome' section, 2013 survey of 'Hear and Treat' callers, Trust B**

| Respondent | Scores | | | | Weight |
|------------|----------------------|-------------|--------------|-------------|--------|
|            | Understand Advice | Explain Adv | FollowAdvice | ExplainAmb  |        |
| 1 | 1.586 | 1.586 | 1.586 | 1.586 | 1.586 |
| 2 | 0.755 | 0.755 | 0.755 | 0.755 | 0.755 |
| 3 | . | 1.313 | 1.313 | . | 1.313 |

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 four questions that contributed to the 'Outcome' section of the questionnaire.

UnderstandAdvice:     $\dfrac{15.86 + 0}{1.586 + 0.755}$     =     6.755

ExplainAdv:     $\dfrac{7.93 + 7.55 + 0}{1.586 + 0.755 + 1.313}$     =     4.236

FollowAdvice:     $\dfrac{7.93 + 3.775 + 13.13}{1.586 + 0.755 + 1.313}$     =     6.797

ExplainAmb:     $\dfrac{15.86 + 3.775}{1.586 + 0.755}$     =     8.387

**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 'Outcome' section of the survey of 'Hear and Treat' callers would be calculated as:

$(6.755 + 4.236 + 6.797 + 8.387) / 4 = 6.451$

**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 score[1].,
$y_i$ is the trust score
$\theta_0$ 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" (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 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.

---

[1] Calculated using the method in Appendix C.

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 $_{(1-q)}$. 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 $\hat{\phi}$.

<div align="center"><strong>An additive random effects model</strong></div>

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 $\theta_i$ (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:

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

## 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:

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

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.