NHS Patient Survey Programme

2017 survey of women's experiences of maternity care

Guidance for trusts on using survey outputs

Published January 2018

CQC publication

Introduction

The 2017 survey of women's experiences of maternity services is part of the NHS Patient Survey Programme. The programme is co-ordinated by the Patient Survey Co-ordination Centre, based at the Picker Institute Europe, on behalf of the Care Quality Commission (CQC).

The 2017 survey of women's experiences of maternity services involved 130 NHS acute trusts in England. We received responses from 18,426 women a response rate of 37.4%.

Women were eligible for the survey if they had a live birth during February 2017, were aged 16 years or older, gave birth in a hospital, birth centre, maternity unit, or who had a home birth.[1] NHS trusts in England took part in the survey if they had a sufficient number of eligible women that give birth at their NHS trust during the sampling time frame.

As in previous years, most of the questions in the questionnaires were 'closed' questions, where the respondent had to tick the option that corresponded most closely to their experience. However, there is a section for respondents to provide other comments ('free text' comments) at the end of the questionnaire. Trusts are able to use the free text comments though **must** ensure that any publication of the feedback does not breach patient confidentiality.

This document contains guidance for trusts on how to use the survey data produced by CQC for each trust, due for publication on the CQC website on Tuesday 30th January 2018.

Using the scored data

The results for the labour and birth section of the questionnaire will be published on the CQC website and are also available in your benchmark report. We only publish the labour and birth data as the responses that women gave to those questions would definitely have been referring to care received from your acute trust, rather than other providers. Some women may have received their antenatal and / or postnatal care through other providers, and therefore their responses do not refer to the acute trust.

Benchmarked results have been sent to the survey leads at each trust, and the reports will be available on the NHS Surveys website from the publication date (Tuesday 30th January 2018).

The survey data can be found on the CQC website by searching for an organisation from the home page, clicking on the NHS trust name, then selecting the survey under the 'Surveys' tab or from 'Latest patient surveys' on the right hand side.

The presentation of the survey data on the CQC website 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.

We recommend using the analysed data contained in both the benchmark reports and shown on the CQC website to identify areas where your organisation has done well, or where you are aware that performance could be improved if you are ready to commit to making those changes in any publicity.

The survey data has been standardised and scored so that trusts' results can be more fairly compared against others, and are intended to allow trusts to easily identify areas of good

[1] Some trusts with a small number of women delivering in February also included women who gave birth in January 2017. For further details on women excluded from the survey, please see the survey instruction manual.

performance, and where improvements are required. Each NHS trust received scores out of 10 for each question, based on the responses given by their respondents. A higher score is better.

The results for each question that could be scored are shown alongside a category identifying whether the trust's score is 'better', 'about the same' or 'worse' than would be expected, when compared with the results from most other trusts. Detail on the scoring and analysis are available in the technical document, this was sent to trust survey leads alongside their benchmark reports, and will be published on the CQC website alongside the survey results.

The maternity survey attribution exercise

Your NHS trust may have more than one benchmark report for the maternity survey. As well as the main benchmark report covering the labour and birth data, your trust may also have received an antenatal report and/or a postnatal report, depending on whether staff were able to complete the attribution exercise, in which we asked trusts to identify women who were likely to have received their antenatal and postnatal care from the trust. The data in the antenatal and postnatal reports must be used with caution, for the reasons described below.

In total, 126 trusts (out of 130) were able to complete the attribution exercise and identify women that were likely to have received their antenatal care and / or postnatal care from their trust, based on their electronic records, or if not available home address (via partial postcodes). This information was used to identify the respondents who were likely to have been referring to the acute trust when responding to the antenatal and postnatal care sections of the questionnaire. Scored results were then produced based only on those respondents, and reports produced for antenatal and postnatal care.

Of these 126 trusts, 126 trusts provided **postnatal** attribution data, and 125 trusts provided **antenatal** data. Three trusts had their postnatal data suppressed due to low numbers of respondents (less than 30), and 2 trusts had their antenatal data suppressed due to a low number of respondents. Therefore a total of 123 trusts received a postnatal benchmark report and the same number of trusts received an antenatal benchmark report.

The data for the antenatal and postnatal sections cannot be considered as statistically robust as the data for labour and birth, for several reasons:

1. Although the value of the data is improved when looking at individual trust performance, due to the more accurate attribution of responses to provider, the lack of complete coverage across all trusts means that we cannot fairly say that one trust is 'better' or 'worse' than all others. Hence trusts are only identified as being 'better' or 'worse' within the subset of trusts that completed the attribution exercise. We cannot say that the subset of trusts is representative of all trusts, and so it is not a true benchmark for performance across England.
2. The attribution was based on the location of respondents. There were no means available to identify women who had received care from a different provider for other reasons, such as due to requiring specialist care, or having moved house during pregnancy. So although the attribution exercise improved the data to a considerable degree, it may remain that some respondents are included in the data despite having received care from another provider.
3. The NHS trusts completed the attribution themselves, and due to the limitations of the process the co-ordination centre were unable to verify the accuracy of the exercise. This means we cannot be certain about the reliability of the attribution of the data, as there were limited opportunities to check for errors.

The antenatal and postnatal survey data from the trusts that completed the attribution exercise will be shared with those trusts. The data will be considered by the Care Quality Commission (CQC) to inform its intelligence model and will be shared with CQC inspectors. The reports will be published on the Patient Survey Co-ordination Centre website**,** but not on the CQC website for the reasons described above.

Those trusts with antenatal and postnatal benchmark reports should bear in mind the above caveats when viewing their data.

Using the percentage data

Your trust survey lead(s) will also have been sent a Word document that shows the survey results for your trust as the percentage of respondents to each response option, for each question in the survey, alongside the total number of respondents for that question. This data will not be published by CQC, as it does not provide a fair comparison across trusts in terms of performance – due to it not being weighted according to respondent characteristics. It is intended for descriptive purposes only, and it is recommended that you use the scored benchmark report data to report comparative performance.

What to consider when reporting survey data

The scored data described above could be reported in press releases, posters, staff newsletters, and other communications, to give an indication of how your trust has performed on the survey when compared with most other trusts. We recommend that you don't mix different types of data, such as the percentages of respondents and scores, as this can give a misleading picture to the public. The scores have the advantage of taking the full set of responses into account within just one figure (e.g. they cover 'Yes, definitely', 'Yes, to some extent', and 'No' responses). The scored data has also been standardised to allow for fairer comparison of results between trusts with different demographic profiles (we have adjusted the data to take into account differences in the characteristics of respondents that might influence their answers irrespective of the care they received).

Using the free-text comments

At the end of the questionnaire, respondents were able to add written comments about anything that was particularly good about their care, anything that could be improved and any other comments.

These comments are useful as they can add some insight into how the trust might be able to improve. For example, if your trust performed poorly in relation to an issue in the questionnaire, the comments might offer an explanation as to why this might be an issue, and hence, how it might be improved. You may find it useful to 'theme' the comments, for example by grouping comments that relate to communication, staff, and so on. If your trust used an approved contractor to undertake the survey on your behalf, they may do this for you, depending on the contract arrangements that you have in place with them.

Trusts must ensure that any publication of the feedback does not breach patient confidentiality.

QUESTIONS AND ANSWERS

This section answers some of the questions you may have on the trust level results, as provided in the benchmark reports, and on the CQC website.

The Benchmark Reports

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

The coloured bars represent the full range of all trust scores, from the lowest score achieved by a trust to the highest. The orange section in the charts represents the **expected range** for a score for a trust. This is the range within which we would expect a particular trust to score if it performed 'about the same' as most other trusts in the survey. If a score falls above or below the expected range it will be in the 'better' or 'worse' category, represented by green and red areas respectively. The calculation of the expected range takes into account the number of respondents from each trust as well as the scores for all other trusts, and allows us to identify which scores we can confidently say are 'better' or 'worse' than the majority of other trusts (see the technical guidance for more details, available from: **http://www.cqc.org.uk/maternitysurvey** and sent to survey trust leads prior to publication).

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

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

About the Scores

Why are the scores presented out of ten?

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

How are the scores calculated?

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

About the Analysis

What is the 'expected range'?

The better / about the same / worse categories are based on a statistic called the 'expected' range that is calculated for each question for each trust. This is the range within which we would expect a particular trust to score if it performed about the same as most other trusts in the survey. The range takes into account the number of respondents from each trust as well as the distribution of scores for all other trusts, and allows us to identify which scores we can confidently say are 'better' or 'worse' than the majority of other trusts. Analysing the survey information in such a way allows for fairer conclusions to be made in terms of each trust's performance. This approach presents the findings in a way that takes account of all necessary factors, yet is presented in a simple manner.

It is the same analysis technique as applied to the risk ratings in the Intelligent Monitoring system, and is based on identifying outliers through the use of adjusted Z scores. More detail on this is available in the technical document.

Why are the percentage results for all trusts not provided?

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

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

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

C18: Thinking about your **care during labour and birth**, were you involved enough in decisions about your care?

	Trust A	**Trust B**
Yes, always	59%	59%
Yes, sometimes	10%	39%
No	31%	2%

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

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

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

The reason for 'standardising' data is that we know that the views of a respondent can reflect not only their experience of NHS services, but can also relate to certain demographic characteristics, such as their age. Women's experiences may also vary if they have previously had a child (we refer to this as parity). For example, older respondents tend to report more positive experiences than younger respondents. Because the mix of women varies across trusts (for example, one trust may have more older mothers), 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 women. To account for this we 'standardise' the data. Standardising data adjusts for these differences and enables the results for trusts with different population profiles to be compared more fairly than could be achieved using non-standardised data.

Why are there no confidence intervals surrounding the score?

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

Understanding the Data

Why do most trusts appear to be performing 'about the same'?

The expected range is a conservative statistic. It accounts for the possibility that there is variation across trusts for other reasons, aside from differences in trust performance. There may be significant variation between trusts due to certain factors that are not within the trusts' control. The technique used takes this into account, and so if a trust is found to be performing 'better' or 'worse' compared with most other trusts that took part in the survey, you can be really very confident that this is the case and it is extremely unlikely to have occurred by chance.

Even though your trust may appear to be performing 'about the same' compared to most other trusts, the results should still be useful to you locally, for example you may want to:

- Make comparisons to the results from previous surveys to look for questions where you have improved or declined.
- Compare your results with those of other similar trusts.
- Look at your results by different sub groups to understand their different experiences, for example, by age, parity, ethnic group, etc.
- Identify particular areas you may wish to improve on ahead of the next survey
- Undertake follow up activity with women such as interviews, workshops or focus groups to get more in depth information into areas in which you would like to improve.

Please remember that for points 1-3 above, to do this accurately you should undertake an appropriate **significance test.**

The survey instruction manual provides more information on making use of survey data.

Why does the number of trusts performing 'better' or 'worse' at each question vary?

It is important to be aware that the ranges of performance on different questions varies and this has an influence on how much a trust needs to differ from the average by, in order to be considered 'better' or 'worse' than the average. This means that the number of trusts performing 'better' or 'worse' at each question will vary.

Why has no trust come out as performing better or worse for a particular question?

This can occur in the analysis of the data and is an acceptable consequence of the statistical technique that is used. The size of the expected range is constructed by considering how different all trust scores are across the range, as well as the confidence we can have in that particular trust's score (by looking at the number of respondents to that question). In some cases, this will lead to such a wide margin of error that the 'expected range' will be very wide, and hence will also cover the highest and / or lowest scoring trusts for that question.

Is the lowest scoring trust the worst trust in the country, for each question? And likewise the highest scoring trust the best?

If a trust is in the 'better' or 'worst' category this mean that they are performing either better or worse compared with *most other trusts* that took part in the survey. However, a trust is not necessarily *the best*, or *the wors*t, and this could not be determined without undertaking an appropriate significance test.

If you took the scores and ordered them by size, you would most likely find that the highest and lowest ones would change if you ran the survey again. This is because the scores are estimates – we have only had questionnaires from some women who used maternity services (those giving birth in February), not all. If another sample of women were surveyed, and you put the scores in order again, you would find that there would probably be a different trust at the top and at the bottom. By analysing the data the way we have, we can say which trusts are likely to always be 'better' and those that will always be 'worse', so they should be looked at as a group of 'better'

trusts, and ‘worse’ trusts, rather than in order of scores. This is the fairest way to present the data as it means that individual trusts are not pulled out as the very ‘best’ or very ‘worst’, when that may not be the case and it may be that if all women were surveyed, different trusts would be shown to be the very ‘best’ or ‘worst’.

The score for one of my questions has gone up but is categorised as ‘about the same’ yet in the 2015 survey we were ‘better’?

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 next time the survey was carried out, 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 useful to look at actual changes in scores over time.

We are categorised as ‘about the same’ for a question yet a trust with a slightly lower score than us is categorised as ‘better’. Why is this?

The ‘expected range’ calculation takes into account the number of respondents from each trust as well as the distribution of scores for all other trusts, and allows us to identify which scores we can confidently say are 'better' or 'worse' than the majority of other trusts. As set out above the expected range is a conservative statistic: it accounts for the possibility that there is variation across trusts for other reasons, aside from differences in trust performance. There may be significant variation between trusts due to certain factors that are not within the trusts’ control. The technique used takes this into account. It is likely that your trust came out as ‘about the same’ because your trust had fewer respondents to the question which creates a greater degree of uncertainty around the result. The trust with the lower score would likely have had more respondents to the question, and so their expected range would have been narrower.

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

This can happen because the calculation of the section scores is a separate calculation and not an average of all questions that make up a particular section. If this has occurred, it is likely that your trust scored very lowly or even on the threshold for all or most of the questions that are in a section.

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

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

Why do the results and / or number of respondents provided by CQC differ from those provided to me by our approved contractor?

CQC do not see the reports provided to you by your approved contractor and therefore cannot comment on these. You should raise any queries directly with your approved contractor. However, likely reasons for any discrepancies are:

- The approved contractor may have cleaned the data differently to CQC. In particular, CQC remove respondents from the base of a question that do not analyse the performance of a trust - we refer to these as ‘nonspecific responses’, such as ‘don’t know or can’t remember’. For more information please see the data cleaning guide.
- Trust level data published by CQC has been ‘standardised’ by age and parity to enable fairer comparisons between the results of trusts which may have different population

profiles. Approved Contractors may not have done this or may have applied a different standardisation. To be able to standardise the data, information is needed on both age and parity, if either of these pieces of information are missing, or not able to be determined, the respondent must be dropped from the analysis as it is not possible to apply a weight.

- CQC analyses trust level data by scoring (and standardising) the responses to each question. Each response option that evaluates performance is scored on a scale of 0-10. Approved Contractors may have analysed and / or scored the data in a different way.
- The Approved Contractor will not be able to make comparisons against all trusts that took part in the survey, only against those that commissioned them. Therefore any 'national' results they publish will not be based on all trusts and any thresholds they calculate may be different.

Comparing Results

Why is statistical significance relevant?

Survey scores are estimates – we have only received questionnaires from some women who gave birth during the sampling period, not all, as the survey uses a sample of women from a chosen month (February, possibly also January) and some choose not to respond. If another sample of women were surveyed, you may find the results would change slightly. This is why it is important to test results for statistical significance.

A statistically significant difference means that the change in the results is very unlikely to have occurred by chance. Without significance testing you cannot be sure that a difference between two results would still be different if you repeated the survey again. If a result is not significant then you cannot be sure of its accuracy. If a significant difference is present then it is likely that it is a true difference, and if the survey was repeated again that you would see the same outcome.

How can I make comparisons to previous years' survey data, or to other trusts?

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

The labour and birth benchmark report for each trust includes a comparison to the 2015 survey scores, where possible, and indicates whether the change is statistically significant. Please note that comparative data is not shown for the sections as the questions contained in each section can change year on year.

The previously published results for the 2007, 2010, 2013 and 2015 maternity surveys are available on the NHS Surveys website. However, given the issue around attribution of survey responses to providers, there is a limit to how accurately the maternity survey data can be compared across years. We have included the 2015 comparisons within the labour and birth benchmark reports this year as they are the only reliable comparisons that can be made – in both surveys, women were definitely referring to care received from your acute trust. All other survey results, for other questions, would have included women who had received their care elsewhere, and so were reporting on the performance on other organisations. This will hinder the accuracy of comparisons across years as the proportions of women referring to your organisation may have varied, and we cannot identify or control for this.

For advice on making accurate comparisons you may like to speak to someone within your trust with statistical expertise, or your approved contractor (if used) should be able to advise on this. You can also contact the survey team on **patient.survey@cqc.org.uk**.

Which trusts are performing best / worst?

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

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

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

We have also published a separate report identifying outlier trusts and this is available on the CQC website.

Why can't I sort the scores for all trusts and rank the trusts in order of performance?

It is not appropriate to sort the scores:

1) Firstly, due to the analysis technique applied, where the number of respondents is taken into account, it is possible that one trust may score higher than another - though the higher scoring trust is classed as 'about the same' and the second, lower scoring, trust is put into the 'better' category. This may occur if the second trust has a considerably larger number of respondents, as it will be assumed that their score is more reliable, and hence more likely always to be high.

2) Secondly, the statistical technique does not measure how different individual trust scores are from one another (whether statistically significant), and so it would be too simple to attempt to sort by scores alone, without running more analysis on the data. The banding technique used is helpful in identifying which trusts are likely always to be in the 'better', 'worse', or 'about the same' category, no matter how many surveys are sent out.

Can I see results for my local hospital / maternity unit etc.?

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

Further Information

The full national results for the 2017 survey are on the CQC website, together with an A to Z list to view the results for each trusts labour and birth questions, and the technical document outlining the methodology and the scoring applied to each question:
www.cqc.org.uk/maternitysurvey

For the trusts who compiled attribution data, the reports for antenatal and postnatal care are available on the NHS surveys website, along with the labour and birth reports for all trusts, at:
www.nhssurveys.org/surveys

The results for the 2007, 2010, 2013 and 2015 surveys can be found on the NHS surveys website at:
www.nhssurveys.org/surveys/299

Full details of the methodology for the survey can be found at:
http://www.nhssurveys.org/surveys/1055

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

Feedback

We welcome all feedback on the findings of the survey and the way we have reported the results – particularly from people using services, their representatives, and those providing services. If you have any views, comments or suggestions on how we could improve this publication, please patient.survey@cqc.org.uk

We will review your feedback and use it as appropriate to improve the statistics that we publish across the NHS Patient Survey Programme.