

NHS Patient Survey Programme

Survey scoring method

November 2015

Presentation of results

The NHS Patient Survey Programme allows patients to feedback on the quality of care they receive from NHS organisations. Survey questions have been developed to allow patients to tell us about their experiences, focusing on areas that we have been told are important by patients and other stakeholders. These questions are primarily of multiple-response format.

Results are presented by CQC in two main ways. For England level results, results are presented as percentages showing the total proportion of patients choosing each response option; while trust level results (i.e. how well each NHS trust performed on the survey) are presented as scores¹. More information about the different outputs provided by CQC is available in the ['Finding survey data'](#) document.

This document sets out the rationale for using scores to present trust level results.

How does scoring work?

All survey questions have more than one response option which means interpreting results for several different answer options simultaneously can be challenging, particularly when comparing trusts against one another or when making comparisons with data from an earlier survey. For this reason, some form of response 'weighting' or scoring is beneficial to provide a single summary result for each question in each trust. The NHS Survey Programme uses what is known as a 'partial credit scoring' system (where partial credit is given for partial 'success').

When the results for surveys in the NHS patient programme are produced, a mean score is calculated for questions that evaluate the quality of care provided. A question is considered to be evaluative if it assesses patient or service users experience of care and will help the trust identify areas for service improvement. The question and responses have to be attributable to the trust rather than being the responsibility of another provider of care.

It is not possible to assign a score to all questions. This is because not all of the questions evaluate the quality of care. The aim of the scoring model is to enable organisational performance on a survey question to be summarised readily and compared across organisations.

Where scoring is applied, the most positive answer option is 10 and the least positive is 0. Intermediate answer options are scored with intermediate values (for example, the middle of three options would be scored 5). The scoring model allows data relating to a question's multiple response options to be summarised by a single number. Scored questionnaires are always published on www.nhssurveys.org

¹ Trusts also have access to their own unweighted percentage results to see how many patients gave each response, but this data is not suitable for comparison purposes.

An example of the application of scoring and the ensuing trust score calculation is as follows².

Did you have confidence and trust in the doctors treating you? (Trust 1)

Response option	Number of responses	Percentage of respondents	Score assigned	Score calculation
Yes, definitely	300	75%	10	300x10=3000
Yes, to some extent	50	13%	5	50x5=250
No	50	13%	0	50x0=0
TOTAL	400			3250/400= 8.13

Did you have confidence and trust in the doctors treating you? (Trust 2)

Response option	Number of responses	Percentage of respondents	Score assigned	Score calculation
Yes, definitely	210	53%	10	210x10=2100
Yes, to some extent	190	48%	5	190x5=950
No	0	0%	0	0x0=0
TOTAL	400			3050/400= 7.63

Rationale for applying scoring to questions

The scoring method used within the survey programme currently, was initially developed by the Department of Health (DH) and was applied to the Adult Inpatient Survey before the NHS Patient Survey Programme was handed over to the regulator in 2003.³ The 2002 survey published by DH published mean 'ratings' (scores) for each trust.

However, in 2002 a pilot survey was run in advance of the national survey, and results were presented using a dichotomous 'problem score' (scores actually being percentages here) approach which was intended to indicate the presence or absence of a problem (see below). The problem scores on individual questions were then summed into seven 'dimension scores' (see the ['Development and pilot testing of questionnaires for use in the acute NHS trust inpatient survey programme'](#) report for further information.

² Please note, this is a simplified version of the scoring system that is applied to the national surveys. The example does not include standardisation to the respondent population, whereby responses are standardised by the age and gender (and other relevant characteristics such as route of admission) to enable more fair comparisons between trusts with different patient profiles.

³ The programme is currently run by CQC, however our predecessor organisation, the Commission for Health Improvement (CHI) first initiated the programme.

Examples of questions from 2002 Adult Inpatient Pilot Questionnaire showing derivation of problem scores

Black circles indicate responses coded as a 'problem'.

When you had important questions to ask a doctor, did you get answers you could understand?

- 1 Yes, always ☺
- 2 Yes, sometimes ●
- 3 No ●
- 4 I had no need to ask ☺

However the problem score approach was discounted before use in the NHS Patient Survey Programme, and the current 'partial credit' method adopted, whereby response options for a question are ordered from least to most positive in their evaluation of experience. The benefit of this approach was that it recognised more intermediate experience and did not rely on knowing which responses were being used to identify a 'problem'. It is a method frequently used in health assessment research.

The 'partial credit' system was deemed superior at this stage also, because the problem score approach can be considered to be especially beneficial when one wants to focus on quality improvement. Using a problem score, anything other than the best response is deemed a problem and would merit improving. The focus is on excellence and anything else is poor. Owing in part to this perceived negativity, it was considered in 2002 that a problem score method was not necessarily appropriate for calculating performance indicators, (with data from the survey being used to construct performance indicators for DH contributing to ['NHS Performance Ratings'](#) in 2002). It was considered unfair to make comparisons between Trusts without taking into account the different degrees of 'problem' reported by patients – in other words, without taking into account the middle responses or 'to some extent' responses. The partial credit scoring system was instead used and allowed for the coding scale to be adjusted according to the range of response options for each question. The increased 'sensitivity' of the approach made it the method of choice.

Both the partial credit scoring model and the 'problem score' approach have been shown to perform with higher levels of reliability than alternative models (Sizmur, 2014⁴), for example a 'bottom box' approach whereby the least positive (worst)

⁴ <http://www.pickereurope.org/wp-content/uploads/2014/10/The-reliability-of-trust-level-survey-scores-...pdf>

responses are focused upon. With both partial credit scoring and problem scores broadly equitable in terms of reliability and allowing discrimination between trusts, it falls then to consider other benefits of the approach currently employed to analyse trust level survey results. Namely that the partial credit method allows all response options to be taken into account in results (rather than just extreme ones).

The benefits and drawbacks of partial credit scoring

In summary, the 'partial credit' approach has a number of benefits and drawbacks:

Benefits

- Relatively straightforward to apply.
- Puts all evaluative questions (regardless of number of answer options) onto the same scale.
- Reflects the intuitive notion that higher score = better experience.
- Recognises partial achievement of valued experience – more sophisticated than 'right or wrong'.
- Reflects the underlying continuum of experience that becomes segmented during question-writing.
- In principle (and ignoring measurement error) zero represents absence of the valued experience.
- Current system with 10 as the highest question score avoids confusion with percentage results.
- Transferrable to new questions without calibration.

Drawbacks

- Obscures detail that would be revealed in a full breakdown by answer options. On some questions, different organisations can achieve the same score with different combinations of responses (a consistent mediocrity is valued the same as widely variable performance). It is not necessarily clear that this does represent the same level of performance as distribution of responses can lead to the same scores.
- Often assumes equal-interval properties within and across questions, which may be inappropriate.
- Takes no account of the relative salience to patients of different aspects of their experience (a disadvantage if scores from different questions are compared or combined).
- Some find this approach difficult to understand when new to the data – the method requires explanation

At present, the model's simplicity, high degree of trust-level reliability and its capability of accurately discriminating between providers makes it a reliable method of scoring where the benefits currently outweigh the drawbacks. However we are keen to explore usability with those working with the data, and welcome feedback on experiences of working with results presented in this way. If you would like to tell us about your experiences please contact patient.survey@cqc.org.uk