

# Overarching questions for patient surveys: development report for the Care Quality Commission (CQC)

CHRIS GRAHAM & SHEENA MACCORMICK  
NATIONAL PATIENT SURVEY CO-ORDINATION CENTRE

## Contacts

National Patient Survey Co-ordination Centre  
Picker Institute Europe  
Buxton Court  
3 West Way  
Oxford  
OX2 0JB

Tel: 01865 208127  
Fax: 01865 208101  
E-mail: [advice@pickereurope.ac.uk](mailto:advice@pickereurope.ac.uk)  
Website: [www.nhssurveys.org](http://www.nhssurveys.org)

## About the National Patient Survey Co-ordination Centre

The National Patient Survey Co-ordination is managed by the Picker Institute on behalf of the Care Quality Commission (CQC). We are responsible for designing, co-ordinating, and reporting on the findings of surveys of NHS patients and service users conducted as part of the national patient survey programme for England.

## Key personnel

Lucas Daly  
Chris Graham  
Harriet Hay  
Esther Howell  
Caroline Killpack  
Jenny King  
Katherine Körner  
Sheena MacCormick  
Steve Sizmur  
Mark Waters

## Questions and comments

If you have any questions or concerns regarding this document, or if you have any specific queries regarding the submission of data, please contact the Co-ordination Centre:

By e-mail: [advice@pickereurope.ac.uk](mailto:advice@pickereurope.ac.uk)

By phone: 01865 208127

## Table of contents

1	Executive summary.....	4
1.1.	Background.....	4
1.2.	Methods.....	4
1.3.	Results.....	4
1.4.	Recommendations.....	5
2	Introduction.....	6
3	Methods.....	7
3.1.	Cognitive testing.....	9
4	Results.....	10
4.1.	Net Promoter Score.....	11
4.2.	“On reflection, did you get the care that mattered to you?”.....	16
4.3.	“Overall, how would you describe your experience?”.....	21
4.4.	“Overall, how would you rate the outcome of your <experience>?”.....	25
4.5.	Combined outcome/experience question.....	29
4.6.	Best and worst care received.....	30
4.7.	“Overall...”.....	36
5	Discussion.....	41
6	Appendices: interim reports.....	43
7	References.....	77

## Acknowledgments

This report describes a complex and large scale project to evaluate questions across a range of sectors. In presenting this report we gratefully acknowledge the invaluable contributions of:

- The 85 members of the public who shared their time and experiences with us in interviews.
- The Care Quality Commission's Surveys Team, who arranged and undertook all of the interviews that tested the maternity survey with recent mothers.
- All members of the Picker Institute's survey development team, as listed on the previous page, who took part in interviews.

# 1 Executive summary

## 1.1. Background

National surveys provide a powerful and rigorous means of measuring and understanding the experiences of NHS patients and service users. The surveys currently in place typically look at a particular group of patients – for example recent inpatients or outpatients – and include a range of questions about specific reportable events. Generic questions on overall satisfaction are typically avoided because these are often vaguely conceptualised and hard to interpret, but the downside to this is a limited potential to compare overall experiences across settings.

This report describes a project to test and develop a new overarching measure of patient experience for use in a wider range of settings. The aim was to identify a single question that could be meaningfully asked of people from a range of backgrounds to provide consistent information about experiences of different services.

## 1.2. Methods

Cognitive interviews were undertaken with people with recent experience of either being an inpatient, seeing a GP, or using NHS mental health or maternity services. A total of 85 interviews were carried out by the Picker Institute and the Care Quality Commission. Six questions were included and iteratively developed through three rounds of testing, with the order of the questions rotated between rounds to counterbalance for potential order effects. Questions were tested within standard questionnaires from the national patient survey programme to provide appropriate context, and interviews typically lasted between 45 minutes and one hour.

## 1.3. Results

The cognitive testing process raised numerous issues and all items were revised to at least some extent during the process. Results are described in detail throughout this document, with issues raised typically focussing on interviewees' comprehension, interpretation, evaluation, and response to questions. Of the six questions tested:

- We do not recommend the Net Promoter Score for NHS use. Interviewees reacted badly to the concept of 'recommendation', particularly in the mental health setting, and a number of interviewees misunderstood what the question was asking.
- A question on whether people "got the care that mattered to [them]" was generally answered with ease by interviewees. However, there was great variance in how people interpreted and evaluated the question, limiting the value of the item for service improvement or performance assessment.
- A question on overall descriptions of experiences worked reasonably well, but there were issues relating to the use of the word 'experience' in the question stem and the sufficiency of the response options.
- A similar question on overall outcomes worked well, particularly when contrasted to a question on experiences. This item was evaluated based on clinical outcomes, which may make it a useful supplement to an overarching measure of experience.
- A two-part approach asking people about good and bad elements of their experience tested poorly and was confusing for many respondents
- A simple overall question with minimal wording and an 11 point Likert scale for response was well understood by interviewees and proved successful as a measure of overall experience.

## 1.4. Recommendations

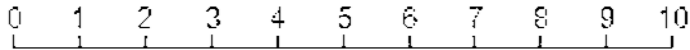
Based on the findings from cognitive testing, we conclude that the following question constitutes a good overarching measure of patient experience:

**Q6.** Overall...

I had a very  
poor  
experience

I had a very  
good  
experience

0 1 2 3 4 5 6 7 8 9 10



We recommend this question for adoption in future national surveys.

As with any comparative measure, care should be taken when implementing this question into different surveys to ensure that differences in results are not attributable to differences in respondent populations. Careful statistical analysis of quantitative data will prove valuable in making use of the item in the future. This should include comparisons to other questions on patient experience to examine the extent to which this overarching measure correlates with more specific reports of experiences. Additionally, analysis should investigate the sensitivity of the measure by looking at whether it discriminates effectively between differently performing organisations.

## 2 Introduction

The NHS national patient survey programme for England includes a range of surveys investigating the experiences of NHS patients and service users. These surveys cover various settings, from community care to acute hospitals and specialist mental health services.

Whilst surveys in different settings have typically used similar approaches, the questions included for surveys of each sector tend to differ. This is largely because the nature of people's interactions with different types of service will vary. For example, it would be inappropriate to ask people who had recently attended an emergency department about their experiences of psychological therapies, because the vast majority of respondents would have no relevant experiences to describe. In spite of this, there are some commonalities in the ways in which people should expect to be treated by different services: for example, most surveys include questions to establish whether healthcare professionals treated patients or service users with respect and dignity.

The benefit of asking similar questions across surveys is that this should make it possible to compare people's experiences across different kinds of services<sup>1</sup>. To support choice and to provide accessible patient and public information, it may be particularly useful for these comparisons to cover global or overarching views of care: that is, indications of the overall quality of people's self-reported experiences of health services.

This document describes work undertaken to test and develop an overarching indicator of patient experience suitable for use in a range of settings. This work was undertaken in November and December 2011 by the Picker Institute and the Care Quality Commission (CQC) and involved a considerable number of cognitive interviews in a range of settings.

---

<sup>1</sup> Notwithstanding the fact that the case-mix of services may differ: this is covered later in section 5.

## 3 Methods

The aim of this project was to test and refine a small number of ‘overarching’ questions in order to identify a single question which would be expected to perform well in a wide range of survey contexts. A challenge here is the sheer breadth of services offered by the NHS: it would clearly be impractical to cognitively test as a set of questions in every service setting in which they may be used. Instead, we sought to test the questions within a diverse range of settings that might be assumed to represent an extreme range of pathways and experiences: we hypothesise that if a question could work consistently across these test areas then it would also be expected to work in other settings that bore similarity to at least one of the tested groups. The patient groups included in the research, and the criteria used to identify participants, are shown below in table 1.

**Table 1: Patient groups and selection criteria**

Group	Criteria
Inpatients:	Participant to have spent 1 night or longer in hospital in the last 12 months
GP users:	Participant to have visited their GP on their own behalf in the last 12 months
Mental Health Service users:	Participant to have used NHS Mental Health services in the last 12 months
Maternity services users:	Participant to have given birth in the last 12 months

In order to put the overarching questions in context and so that respondents had reflected on their NHS experience prior to answering, the overarching questions were inserted into abridged versions of the most current Picker Institute national survey questionnaires, normally towards the end of the questionnaire.

Three rounds of interviews were conducted with each round containing approximately 8 interviews per patient group (with the exception of Maternity where slightly fewer interviews were required). After each round there was a one week pause where findings were reviewed and the overarching questions revised or replaced where necessary. An interim report was written on the completion of each round, and the order of the overarching questions was rotated for the subsequent round to counterbalance for potential order effects.

The dates for each round of interviews were as follows:

- Round 1: 17<sup>th</sup> – 28<sup>th</sup> October 2011
- Round 2: 7<sup>th</sup> – 18<sup>th</sup> November 2011
- Round 3: 28<sup>th</sup> November – 9<sup>th</sup> December 2011

The cognitive interviews were conducted by staff from the Picker Institute and Care Quality Commission. Interviews took place in Oxford (GP, Inpatient and Mental Health) and London (Maternity). A recruitment company was used to recruit the majority of respondents in Oxford, although, due to the more complex nature of the Mental Health interviews, participants were also recruited via the local press. Maternity respondents were mainly recruited from a Children's Centre in central London, with one mother recruited via Netmums.com.

Interviews were conducted face-to-face and either took place in the Picker Institute offices or, where more convenient, in a location closer to the participant's home or work address. A small number of respondents were interviewed at home.

Participants were given a brief background to the study and were then asked to complete a copy of the relevant questionnaire. They did this by reading each question aloud and, where applicable, explaining the reasons for their response. Particular care was taken to understand how participants responded to the overarching questions. Interviews lasted on average 45 minutes – 1 hour.

No fixed quota was specified for this project but the aim was to ensure a good mix by age, gender and ethnic background. Details of interviews completed by demographic group are provided below in tables 2 and 3.

**Table 2: Interviews conducted by age and gender of interviewees**

Gender	Age group	People who have visited their GP	Inpatient	Community Mental Health service users	Maternity	Total
Female	16-35	3	3	7	9	22
	36-50	2	1	3	2	8
	51-65	7	4	3	0	14
	66+	1	3	0	0	4
Male	16-35	4	5	5	.	14
	36-50	1	2	1	.	4
	51-65	6	4	4	.	14
	66+	2	2	0	.	4
<b>Total</b>		<b>26</b>	<b>24</b>	<b>23</b>	<b>12*</b>	<b>85*</b>

\*12 women were interviewed with regard to their maternity experience, but 1 interview had to be terminated before demographic information could be collected (all overarching questions were covered in the interview).

**Table 3: Interviews conducted by ethnicity of interviewees**

Ethnic Background	People who have visited their GP	Inpatient	Community Mental Health service users	Maternity	Total
White	23	23	21	7	74 (87.1%)
Asian/Asian British	2	1	1	1	5 (5.9%)
Black/Black British	1	0	1	2	4 (4.7%)
Other Ethnic Group	0	0	0	1	1 (1.2%)
<b>Total</b>	<b>26</b>	<b>24</b>	<b>23</b>	<b>12*</b>	<b>85*</b>

\*12 women were interviewed with regard to their maternity experience, but 1 interview had to be terminated before demographic information could be collected (all overarching questions were covered in the interview).



### 3.1. Cognitive testing

We used cognitive interviewing to test the questions. Participants were asked to complete the questionnaire whilst observed by a researcher to confirm that their understanding of the questions was as intended and therefore to demonstrate, non-statistically, the content and construct validity of the items. In doing this, we considered the cognitive process of responding in terms of the model described by Tourangeau (1984)<sup>i</sup>, seeking to establish consistency in:

- **Comprehension** – people understand what the question is asking in a consistent way that matches the intended research question.
- **Retrieval** – people are able to retrieve from memory the information necessary to evaluate their response to the question.
- **Evaluation** – people are able to use retrieved information to evaluate the question meaningfully, and do this in an unbiased manner (eg, not simply acquiescing or providing socially desirable responses).
- **Response** – people are able to match their evaluation to one of the available responses in a meaningful and appropriate way; the response selected adequately reflects the person's experience.

Cognitive testing of the questionnaire followed an iterative process, with three rounds of testing undertaken. After each round the research team reviewed findings and made changes to the questionnaire as required.

## 4 Results

As described above, interviews were conducted in three rounds, each with approximately equal numbers per group, and items were reviewed and revised between rounds. It should be noted that the number of interviews undertaken within any of the three main groups is high and the total number of interviews exceptionally so: this gives us confidence that the majority of issues associated with items should have been identified.

Because of the iterative design process used, multiple versions of some questions were tested, whilst others were substantially changed between rounds. Rather than describe each round of testing separately – which would be redundant given that interim reports of each round have already been prepared and are included here as appendices – in this section we seek to draw out key themes relating to each of the questions. For convenience and for clarity questions are described separately, although in some cases we found that issues overlapped between different items: this is identified where relevant.

Throughout this section specific examples of comments from interviewees are used to illustrate key themes. To add context, each comment is prefixed with a two or three character code to indicate the patient group the comment was from. These codes are as follows:

- **GP** General practice survey
- **IP** Inpatient survey
- **MAT** Maternity survey
- **MH** Community mental health survey



- Objection
- Scale and response
- Other issues

## The concept of 'recommendation'

Many interviewees had problems with the concept and terminology of recommendation. Broadly speaking, these problems fell into two main categories: misattribution and objection.

### Misattribution

Some interviewees in all rounds of testing did not understand that they were being asked to recommend a specific service and provider (for example, Oxford's John Radcliffe Hospital for inpatient stays, or a mental health trust for mental health issues). Instead, a sizeable number of respondents interpreted the question as asking whether they would recommend having treatment for a condition – irrespective of the provider:

- **MH** One interviewee immediately circled the option 10 – but told us that this simply reflected the fact that they would always encourage people with mental health issues to seek professional help.
- **IP** One person told us that the question was odd because “if someone was ill you'd say 'go to hospital'”. Again, the question was interpreted to be about advising people to access healthcare when it was needed – not about choosing one particular provider over another.
- **GP** An interviewee who had expressed serious concerns throughout the course of the interview regarding their treatment by a practice nurse nevertheless selected option 10 for this item, believing it to be about seeking care when it was needed.
- **MH** Two interviewees in the final round of testing felt that they would always recommend that people with mental health problems contact mental health services. One ticked 10 and said that as their treatment was free “it would be silly not to recommend it”, in spite of the fact that their own experiences had been distinctly unsatisfactory. Another ticked 9, explaining that it was “better to do something than nothing”: they would have selected 10 but that “some people may be reluctant” to seek help for a mental health condition. Again, this person's positive response was not in keeping with their own experiences.
- **MAT** An interviewee was unsure what the question was asking about and said that it was “weird”. She concluded that it was asking if she would recommend a water birth, as she had had, rather than the hospital where this took place.

Other interviewees understood that they were expected to answer with regard to their particular provider, but raised issues around the definition of this:

- **GP** One interviewee noted that they would be “more likely to recommend [their] GP than the practice” and was consequently unsure of how to answer.
- **GP** An interviewee receiving regular treatment related to alcohol dependency stated that they would recommend their GP for people needing similar treatment, but not for those with other conditions. They were therefore unsure how to answer the question, even though this interview was conducted in round two when the suffix “...if they needed similar care or treatment” had been added to the question stem.

## Objections

Objections to the term 'recommend' were raised in all rounds, and ranged from people reacting with amusement to angrily deriding it. Some respondents went so far as to flatly refuse to answer the question: although the number who did this is relatively small, the importance of this should not be underestimated. Except in the case of sensitive questions (eg about sexuality or criminal behaviour) our experience is that it is very unusual for people to actively refuse questions in cognitive interviews where a researcher is present. In part this will be due to socially desirable responding: interviewees are likely to want to please researchers by showing that they are able to answer all of the questions put to them. Since this bias is related to the presence of an interviewer it should be greatly reduced in self-completion settings, and we would therefore anticipate more frequent refusals of this question in postal surveys.

- **IP** One interviewee refused to answer the question, stating that "I don't go around recommending hospitals".
- **IP** In the first round of testing, prior to the addition of the "...if they needed similar care or treatment" suffix, one person who had had an emergency admission to hospital told us that they "wouldn't recommend hurting your hip". They refused to answer the question because they felt that the idea of 'recommendation' implies that people have a choice about the services they use, which is not the case for emergency admissions.
- **IP** In the second round of testing an interviewee refused to answer the question, stating that as they had not seen every doctor in the hospital it would be "stupid" to recommend the hospital based only on the doctors that they had seen.

A number of interviewees did answer the question in spite of raising concerns about the concept of recommendation

- **IP** Two further interviewees used very similar phrases to describe the question. One said that "I simply don't recommend hospitals" before ticking option 8, whilst another said that "I don't recommend hospitals... it's ridiculous" – but nevertheless circled option 9.
- **MAT** One interviewee noted that "you would not recommend NHS care because you don't pay for it".
- **IP** An interviewee in round three seemed confused by the question, asking "why would I recommend bladder cancer?". They nevertheless circled option 8.
- **MAT** One interviewee considered the concept of "recommending the NHS" as confusing because the NHS is "just there for everyone". They eventually interpreted the question as being about the choice between NHS and private care, rather than their particular hospital.

## Scale and response

A number of interviewees gave unusable responses to the NPS, either by multicoding the question (selecting more than one response option) or by writing in a non integer value (eg 8.5). This latter type of error in particular occurred frequently in rounds one and two: consequently we removed the line from beneath the response options prior to round three to try to discourage interviewees from thinking of the responses as representing a continuous rather than an ordinal scale. This change appeared to be successful as for the first time none of the interviews resulted in a non-integer response.

## Other issues

An issue noted in one interview raised concerns about whether the item was an appropriate measure of particular episodes and whether it may be particularly susceptible to bias.

- **IP** An interviewee who had attended the same hospital on numerous occasions over a period of several years told us that their answer was based on all of their hospital

experiences, not just their recent one. Although they had no difficulties concentrating on their recent episode when answering questions, they thought it odd to be asked to recommend the hospital based purely on a single visit.

The concern here is that for some respondents at least it may be difficult to focus in to a particular experience when trying to make a recommendation – particularly if that experience was atypical.

Another issue noted across the full range of interviews was accuracy of the measure: interviewers frequently reported that people's responses to the NPS did not feel consistent with the experience that they had described. For example:

- **IP** A female interviewee who was admitted to hospital whilst pregnant ticked 8 for the item in spite of describing an extremely distressing hospital episode (a minor condition was misdiagnosed as an ectopic pregnancy and she was incorrectly told that she had lost her baby). This person would be classed as 'passive' in a standard analysis of the NPS.
- **MH** One interviewee told us that they "would recommend" the services that they received, but wrote in a score of 6.5 "because nothing's perfect": this person would be classed as a 'detractor' in a standard analysis of the NPS.

Several other interviewees gave responses that similarly failed to measure up to their experiences, although these were often related to other issues. Several such cases are described in relation to earlier themes. These cases arguably support research by Schneider et al (2008)<sup>ix</sup> that the NPS does not provide an accurate indication of actual recommending behaviour.

Finally, some linguistic issues were noted:

- **GP** An interviewee whose first language was not English had difficulty with the question, particularly with interpreting the word "neutral", which they repeatedly read aloud when considering the question.

### 4.1.3. Conclusions

Extensive cognitive testing of the original NPS item and slightly revised versions revealed numerous serious issues relating to the acceptability of the question. As described above, issues were identified in terms of the terminology used, the response scale provided, and the accuracy of measurements from the item.

The key problem with the item is the term 'recommend'<sup>1</sup>, which is fundamental to the concept of the NPS. This prompted serious concerns from respondents in all three rounds of testing. Far too many of the people that we interviewed either misunderstood the question or expressed annoyance at or objection to it, and it was not possible to revise the question to address these concerns without fundamentally changing its design.

Given these issues and academic doubts about the quality and validity of the NPS we are unable to recommend it for use in surveys of NHS patients. Furthermore, the evidence from our testing is that the question would be considered actively objectionable by a not inconsiderable proportion of survey recipients. We therefore conclude that the NPS is not appropriate for use in an NHS setting.

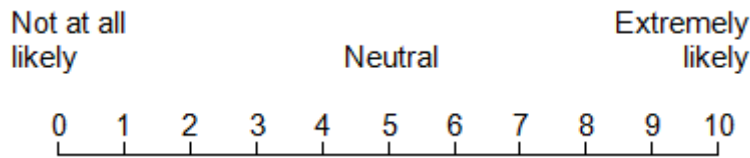
---

<sup>1</sup> It should be noted that two versions of another question on recommendations were also tested: again, these worked poorly because interviewees frequently raised concerns about the term 'recommend'. See section 4.7, below, for further details.

#### 4.1.4. Net Promoter Score – summary of tested versions

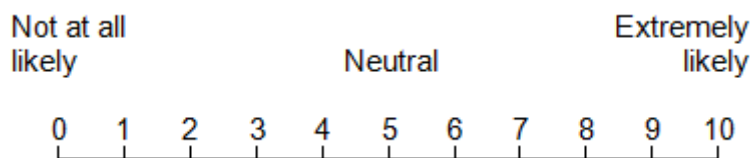
##### Round one – standard Net Promoter Score wording and layout

Q1. How likely is it that you would recommend <this service> to a friend or colleague?



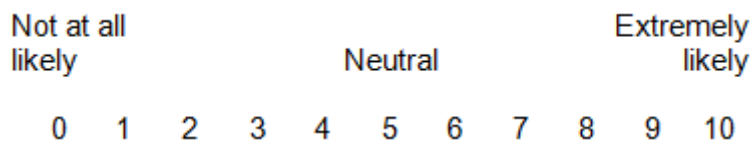
##### Round two – “...if they need similar care or treatment” suffix added

Q1. How likely is it that you would recommend <this service> to a friend or colleague if they needed similar care or treatment?



##### Round three – line removed from beneath responses

Q1. How likely is it that you would recommend <this service> to a friend or colleague if they needed similar care or treatment?



## 4.2. “On reflection, did you get the care that mattered to you?”

### 4.2.1. Background

As with the Net Promoter Score, this question was developed and used prior to the current project. In this case, the question “on reflection, did you get the care that mattered you?”<sup>1</sup> is attributable to NHS North West SHA or North West Inspiration and has originated as part of the Advancing Quality programme<sup>x</sup>. The question uses a similar format and response options to those found in the national patient survey programme:

**Q2.** On reflection, did you get the care that mattered to you?

- 1  At all times
- 2  Most of the time
- 3  Sometimes
- 4  Rarely
- 5  Never
- 6  Do not wish to answer

We understand that at least some trusts use the item, in at least one case as the only question in a discharge survey<sup>xi</sup>. However, we are not aware of any published material describing development or testing of the item, nor are we aware of any publically available data from the question.

Whilst the layout of the question closely resembles items from the national survey, a notable difference is that the question does not ask about a fixed event, experience, or concept. Rather, it is deliberately vague in the sense that it invites respondents to provide their own context – “the care that mattered to *you*”. The rationale for this is, we understand, to truly treat people as individuals and avoid making assumptions about whether healthcare met individuals’ needs and expectations. The trade off for this is that the individual contexts drawn on by respondents will not be known to users of the survey data: as Tourangeau et al (2000) note, “the danger in vagueness is... that some respondents will choose one way to make a vague question precise, whereas others will choose a different way, leading to uninterpretable variability in the responses”<sup>xii</sup>. In other words, two people with identical experiences could give very different answers because they chose a different context, whilst conversely two people with very different experiences could give the same answer. A key question for the cognitive testing was, then, to establish how respondents interpreted the question and what they thought about when answering it.

### 4.2.2. Cognitive testing

Through the course of the interviewing, we tested two versions of the “on reflection...” with a minor change made to remove the “Do not wish to answer” response option after round one. This change was designed to address concerns from interviewees. Tested versions are shown in section 4.2.4, below. In this section of the report, we discuss issues raised across all rounds in terms of the following key themes:

---

<sup>1</sup> Sometimes shortened to just “did you get the care that mattered to you?” or differently prefixed.



- Semantics and comprehension
  - “That mattered to you”
  - “Care”
- The “do not wish to answer” option
- Objections
- Order

## Semantics and comprehension

The main issues to come out of the testing related to people’s comprehension and interpretation of both the question as a whole and the terms used in it. Mostly people were able to interpret the question in a way broadly consistent with the intention, but a minority struggled with what the question as a whole meant:

- **GP** One interviewee said that the question was “fuzzy” and that they took it to be an indirect way of asking if they “had ever been upset”.
- **MH** One interviewee remarked that the question was “too subjective” and that they didn’t understand what it meant.

More specifically, we found issues with the way in which some of the terminology used in the question was interpreted. Whilst almost everyone understood the wording of the question, their specific interpretations of what it meant varied widely. In particular, there were issues with how people decided what mattered to them and what they thought we meant by “care”.

### “That mattered to you”

Throughout the testing we probed people on their responses to the question, asking them what they were thinking of when responding: in other words, what was it that mattered to them? Interviewees gave a predictably wide range of responses, reflecting very different backgrounds, expectations, and preferences. Broadly speaking, these fitted into a number of quite distinct categories:

- **“Everything”**. Some people gave summative answers that took into account all aspects of their experience. Many others focussed on only a single aspect of care.
- **Relational aspects of care**. Some people focussed purely on their relationships with staff or how they treated by healthcare professionals.
- **Transactional aspects of care**. A minority of people stressed the importance of transactional aspects of care such as getting the right tests or having regular antenatal checkups.
- **Clinical outcomes** or ‘getting better’. For example, one interviewee who had been treated for cancer told us that what mattered to them was “being in remission”.
- **Fluid or changing**. Especially in cases of long-term care, some people said that what mattered to them varied at different points in their care.

### “Care”

Whether people identified relational aspects of care or clinical outcomes seemed to depend in part on how they interpreted the word “care”; this seemed to shape what people saw the question as being about or, in some cases, what they saw the questions as *not* being about. Some people believed that “care” only covered the treatment they received, whilst others considered it to be only about relational aspects.

For example:

- **GP** One interviewee told us that they excluded factors like the politeness of staff and the cleanliness of the practice as they did not consider these to be parts of care.

- **GP** Another interviewee said that they were confused as they weren't sure whether they could answer based on having had problems getting through to their practice on the phone: they thought that this probably did not fit the intended definition of care.
- **IP** One interviewee said that "care relates more to people [*ie relationships with staff*] than to treatment".

Issues with the interpretation of the word "care" were most apparent in the first and third rounds of testing, whilst there were few problems in round two when the item appeared first in the set of overarching questions. Interviewers commented that preceding questions may have influenced people's interpretations (see below).

## The "do not wish to answer" option

In the first round several interviewees expressed surprise at the "do not wish to answer" option, stating that it was odd that this was included for this item but not for others. In some cases interviewees seemed concerned at what might be implied by this option, as if it signified something sensitive about the question. None of the people we interviewed selected the item. Because of the concerns about the item and its non-utilisation we removed it for later rounds of testing.

## Objections

No serious objections to the question were raised in the first two rounds of testing. In the third round, however, several interviewees raised concerns about it and stated that they disliked it:

- **GP** One interviewee found the question inappropriate and said that "the care that 'mattered' is irrelevant [because] sometimes you need care that you don't want".
- **MH** One interviewee remarked that the question was "too subjective" and that they didn't understand what it meant.
- **MH** Another interviewee said that the question "seemed pointless".

Because these objections were only raised in the final round of testing when the item was last in the set of overarching questions, interviewers expressed a concern that the question may be particularly susceptible to order effects (see below).

## Order

As noted above relating to the interpretation of "care" and to objections being raised, we were concerned that the item may be susceptible to order effects.

This susceptibility to order effects may be related to the vagueness of the question. Because the item itself does not direct respondents' attention to a particular topic, the details of the topic has to be supplied by the respondent. As many people will not have previously considered what constitutes "the care that mattered" to them, this is a generative process: to respond, one must actively consider and decide what to take into account when answering the question. Consequently for most people the question is cognitively more complex than a question that simply involves reporting something that did or did not happen to them. There is evidence that cognitively more complex questions cause respondents to struggle<sup>xiii</sup> or to rely more on cognitive shortcuts<sup>xiv</sup>: in this instance we hypothesise that this may take the form of using contextualising information from neighbouring questions.

### 4.2.3. Conclusions

Superficially, this question worked well: most interviews were able to answer the question without expressing difficulty and only a few people expressed objections to it. However, problems became evident when probing people on their interpretation of the question.

Interviewees considered a range of different factors when responding to the question. In some cases respondents fixated on the single most important element of their treatment whilst others thought about a number of things. Likewise, some thought about clearly transactional or clinical elements of care, whilst others considered only relational aspects. Most problematically, it was evident that a proportion of respondents made assumptions about what they were *expected* to include, assuming that only some aspects of care would 'count'. Only a minority of people responded to the item in a summative fashion, evaluating their whole experience: many others focussed in on only a single aspect of their care.

As well as this difficulty with interpretation, interviewees were concerned that the item seemed to test differently depending on its placement amongst the six overarching items. Whilst the current work does not include enough quantitative information to adequately test for or demonstrate particular order effects, we are nonetheless of the view that responses to this question are likely to be influenced more by the context of surrounding questions than would be usual. This is potentially a barrier to the comparability of responses, particularly in national surveys where extended question banks are available as an option.

Overall, we believe that the variability in response and concerns around comparability severely limit the usefulness of this question for quality improvement or performance assessment: there is simply too little contextual information provided by the question to interpret it consistently. As such we do not recommend it for use in quantitative surveys.

#### 4.2.4. “On reflection...” – summary of tested versions

##### Round one – standard wording and layout

**Q2.** On reflection, did you get the care that mattered to you?

- 1  At all times
- 2  Most of the time
- 3  Sometimes
- 4  Rarely
- 5  Never
- 6  Do not wish to answer

##### Rounds two and three – “Do not wish to answer” option removed

**Q2.** On reflection, did you get the care that mattered to you?

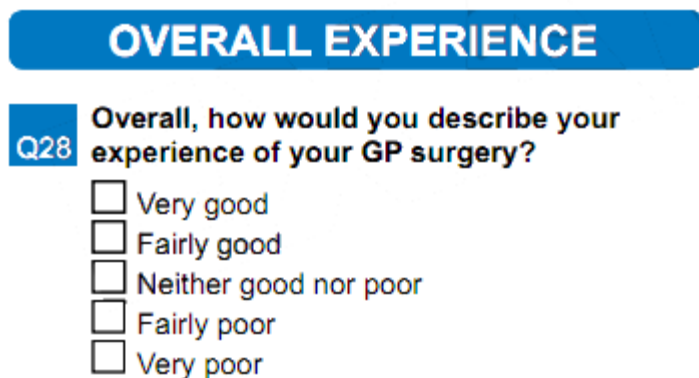
- 1  At all times
- 2  Most of the time
- 3  Sometimes
- 4  Rarely
- 5  Never

### 4.3. “Overall, how would you describe your experience?”

#### 4.3.1. Background

This question has been previously used in the GP Patient Survey (GPPS) conducted by Ipsos MORI on behalf of the Department of Health. Like most other questions in the GPPS it follows a similar layout to surveys in the national programme, and uses a simple direct wording to ask people for an indication of the overall quality of their experience:

Figure 1: GPPS question wording. Source: GPPS January – March 2012



The item is a recent addition to the survey, having been included so far in only one completed round of the survey (July – September 2011: the next round of the survey, January – March 2012, is currently underway). Because results for the July – September 2011 round of the survey have been published, it is possible to get a sense of likely response distributions. PCT level aggregate results are shown in table 4, below:

Table 4: PCT level aggregate response distributions. Source: GPPS

Response	Average	Min	25th pct	75th pct	Max	IQR
Very good	45.7%	32.9%	42.0%	49.7%	58.4%	7.7%
Fairly good	42.1%	32.9%	40.1%	44.4%	47.4%	4.3%
Neither good nor poor	8.6%	5.0%	7.3%	9.8%	13.4%	2.4%
Fairly poor	2.7%	0.9%	2.1%	3.3%	5.4%	1.2%
Very poor	0.8%	<0.5% <sup>1</sup>	0.5%	1.2%	2.7%	0.6%
<b>Good - combined</b>	<b>87.8%</b>	<b>79.2%</b>	<b>86.2%</b>	<b>90.0%</b>	<b>93.4%</b>	<b>3.8%</b>
<b>Poor - combined</b>	<b>3.6%</b>	<b>0.9%</b>	<b>2.7%</b>	<b>4.4%</b>	<b>7.9%</b>	<b>1.7%</b>

<sup>1</sup> The GPPS does not report results where the percentage is less than 0.5%. For this question, 28 out of 151 trusts had less than 0.5% answering very poor, but it is not possible to precisely state the minimum percentage.

From this table, it is clear that the vast majority of respondents describe their practice as either good or very good – and in most cases ‘very good’ is the modal response. There appears to be a floor effect regarding the less positive response options: at most the average proportion ticking either ‘fairly poor’ or ‘very poor’ within a PCT was 7.9%, with a mean of only 3.6%<sup>1</sup>. The results also show limited variation in how people responded, with narrow interquartile ranges for all responses.

There are several important limitations to this rough analysis of the published data. Firstly, results shown above are likely to mask greater variation in practice level results, because there is likely to be a degree of regression to the mean across PCT areas. It is also important to note that the results only cover people’s experiences of GP practices, and that different trends might emerge if the survey covered other types of provider – for example, mental health or acute trusts. Another limitation is that because the question has so far only been used in one completed survey, no longitudinal data is available: we cannot look back across surveys to see whether results have changed over time.

In the GPPS, the question stem includes the suffix “of your GP surgery” to direct respondents to answer about their overall experience of being a patient at the practice. Most surveys in the national programme are slightly different, however, in that they consistently direct people to think about one care episode (eg, ‘your most recent hospital visit’) throughout the questionnaire. This limits the need to provide additional context with the question, so we were generally able to use a shortened wording – the benefit of this being that it should be easier to replicate the question across a range of settings:

**Q3.** Overall, how would you describe your experience?

- 1  Very good
- 2  Fairly good
- 3  Neither good nor poor
- 4  Fairly poor
- 5  Very poor

### 4.3.2. Cognitive testing

Through the course of the interviewing, we tested three versions of this question. In two rounds we used very similar questions, with only a minor change to one response option. These items generally tested well, and with few new issues emerging we chose for the third round to test a merged outcome/experience question.

- Ambiguity of the ‘neither good nor poor’ option
- Interpretation of ‘experience’

For the third round of testing we combined this item with the subsequent one on outcomes. Issues specifically relating to the layout of this merged question are covered in section 4.5 below.

---

<sup>1</sup> It should be noted that these figures marginally underestimate responses of ‘very poor’, because the GPPS does not report results where the percentage is less than 0.5%. For this question, 28 out of 151 trusts had less than 0.5% answering very poor, and we have estimated 0% for these trusts for the purposes of calculation.

## ‘Neither good nor poor’

In the first round of testing the third response option for this item read “neither good nor poor”. This was found to be problematic, with a sizable minority – roughly one third – of all respondents saying that they interpreted it to mean “I have no opinion”, whilst the rest felt that it meant “neutral” or “average”. These interpretations are clearly quite different, creating issues with how the data should be interpreted or used. Consequently, we replaced this response

## Interpretation of ‘experience’

In most cases we found the term ‘experience’ to be well understood and interpreted as intended: interviewees tended to see the question as asking about more than just medical care or clinical outcomes. For example:

- **IP** One interviewee described the question as asking about “more than just [*the hospital*] doing its job” in a medical sense.
- **IP** One person said they considered everything that happened in hospital when evaluating their response.

In spite of this, a number of interviewees had difficulties with the term:

- **MAT** One person said that it was not clear that this question referred to ‘care’ as it did not specify “what experience” was being asked about.
- **MAT** Another recent mother told us that she felt there were two possible interpretations – her experience in the context of her life (ie childbirth and labour being one the most unpleasant experiences she had had) and then the experience of how she was cared for by the service (which was more positive).
- **IP** Two interviewees expressed surprise at the question as they stated that a visit to hospital is never ‘a good experience’. Nonetheless, both gave positive answers to the question.
- **IP** Another inpatient told us that they did not understand what the term “experience” was referring to.
- **MAT** In the final round of testing one interviewee stated that she felt that “outcome” and “experience” “meant the same thing”.

Problems were most pronounced when the question was asked of people’s experiences with antenatal care in the maternity survey. Interviewees tended to think beyond antenatal care and include care received during labour and the birth: others considered a broad range of services including NHS Direct. Although this kind of problem was not evident elsewhere, it nonetheless suggests that additional introductory text may sometimes be needed to help focus respondents on particular services or aspects of care.

### 4.3.3. Conclusions

Overall the question tested well. Concerns about the middle response option were relatively easy to address, and we encountered no further issues with the response scale in later rounds. The only outstanding concern was that some respondents were unsure of or objected to the term “experience”: it is not clear how or whether this could be addressed, but it was not a problem for the vast majority of respondents and we consider this only a minor issue.

Whilst the question performed well in terms of respondents’ understanding and interpretations, a potentially greater issue is the likely distribution of responses in a large scale survey. As data from the GPPS has illustrated, answers to the question are likely to be overwhelmingly positive, potentially limiting the usefulness of the question as an overarching measure.

#### 4.3.4. “Overall, how would you describe your experience?” – summary of tested versions

##### Round one – standard wording and layout

**Q3.** Overall, how would you describe your experience?

- 1  Very good
- 2  Fairly good
- 3  Neither good nor poor
- 4  Fairly poor
- 5  Very poor

##### Round two – middle response option changed to “average”

**Q3.** Overall, how would you describe your experience?

- 1  Very good
- 2  Fairly good
- 3  Average
- 4  Fairly poor
- 5  Very poor

##### Round three – merged into combined outcomes/experiences matrix

**Q3.** Thinking about <the episode>, how would you describe your....

	Very poor	Poor	Average	Good	Very good
<b>Outcome</b>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
<b>Experience</b>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>



## 4.4. “Overall, how would you rate the outcome of your <experience>?”

### 4.4.1. Background

This item, derived from the Outcome and Experience Questionnaire (OEQ) developed by University of Oxford for the Department of Health, is distinct to others tested here in that its focus is on user reported outcomes as opposed to experiences per se. Nonetheless, its layout and response options are very similar to an existing question in the national inpatient survey asking people to rate their overall experience:

**Q4.** Overall, how would you rate the outcome of your <experience: eg ‘visit to hospital’>?

- 1  Excellent
- 2  Very good
- 3  Good
- 4  Fair
- 5  Poor

Because of the difference in approach for the question, our testing focussed on understanding how the term ‘outcome’ was interpreted and whether a positive response here would be a good indication of a good overall experience.

### 4.4.2. Cognitive testing

In the first two rounds of testing we used two subtly modified versions of the question, with the word ‘rate’ being replaced by ‘describe’ and response options being changed in the second round to present the question as more similar to the overall experience question described above.

Key issues identified during the testing related to a number of areas:

- Meaning of ‘outcome’
- Response scale

### Meaning of ‘outcome’

Across all rounds of testing we found that the vast majority of interviewees interpreted “outcome” to refer to the clinical effectiveness of their care or treatment. Many interviewees described this as being about whether they had “gotten better”. Only a minority expressed difficulties with the term in and of itself:

- **MH** One interviewee told us that they didn’t understand the term “outcome”, which they saw as “jargon”. When prompted, they were able to accurately guess what was meant.
- **MH** One interviewee considered the question to be asking for a judgment of “the NHS as a whole” rather than any services they had received. However, this person made similar comments to other questions, suggesting a more fundamental misunderstanding.

Misunderstandings were more common in round two, when the item was presented prior to and apart from the similar question on experience. Several people in this round of testing understood the terminology correctly but assumed based on the rest of the questionnaire content that they were intended to answer in terms of their experiences. In one case, an interviewee made this mistake but went back and changed their response to the 'outcome' question on encountering the subsequent 'experience' question.

Aside from the misunderstandings detailed above, it was clear that most people were able to separate out the outcome of their care from their experience of receiving care. In some cases this was quite striking:

- **IP** One interviewee initially ticked "fair" but then amended their response to "excellent" after noticing the word "outcome" in the question. They explained that their stay in hospital was "average" but that they were "a lot better" as a result of it.
- **IP** One person ticked "excellent" without hesitation, explaining that this was because "I'm just fine now". They noted that there were some serious issues with the care they received – including being misdiagnosed and later being given no information about the outcome of a surgical procedure but said that they did not think these issues were relevant because they had not affected their recovery.

Although most interviewees understood the term 'outcome' well, there were some differences across sectors in how well people were able to answer the question. Few problems were observed in the inpatient questionnaire, where care typically involved a specific surgical or medical intervention to tackle a particular physical health issue. In the mental health and GP surveys, where care is more often on-going, people had more problems:

- **MH** One person said that their treatment was ongoing: they therefore considered it difficult to answer the question because the 'outcome' was yet to be determined.
- **MH** Another interviewee whose care was ongoing said that they interpreted "outcome" as asking "how I'm doing now".
- **MAT** Although the question used in the maternity setting was carefully phrased to try and avoid such issues, one interviewee expressed amusement at being asked to "rate the outcome" of her care because she said that her "baby was the outcome".
- **GP** Several GP patients expressed confusion at this question in the first round of testing because their care was ongoing, sometimes in different settings. One told us that the outcome of their most recent GP appointment was a referral to a hospital consultant, and they were not sure how this should be evaluated.
- **MH** One interviewee raised a slightly different concern. They understood the question well but stated that although their mental health had improved this was not because of the care they had received from the NHS: consequently they were unsure how to answer, but eventually ticked 'average'

## Response scale

Three interviewees raised concerns relating to the response scale used in the first round of testing. Two people noted that the scale was unbalanced, with more positive than negative options, and that they considered this inappropriate. One of these people also noted that the top response of "excellent" may be inappropriate because it required a "very high" standard that might be unrealistic. This later issue was also raised by another interviewee, who said that they could not select "excellent" as they thought this implied "perfection".

### 4.4.3. Conclusions

This question generally tested well, particularly so in the acute inpatient setting. The vast majority of respondents understood the question clearly and had no problems in evaluating their 'outcome' as something quite distinct from their 'experience' – particularly when similarly worded questions on the two areas were presented in close proximity.

The item was more problematic in settings involving long-term care where people may not have had any recent interventions: in these cases changing the item wording to ask people to "describe" rather than "rate" their outcome seemed preferable and raised fewer objections.

Overall, the question worked well as a measure of patients' perceived outcomes. Responses were often quite different from those given to questions on experience, illustrating the point that from the patient's perspective a good outcome will not necessarily be indicative of a good experience. For this reason, we would not recommend using this question alone as an overarching measure, particularly not of *experience* – although there may be benefits to including it within national surveys as a way of collecting summary information on outcomes that may be both complimentary and contrasting to information on experiences. As a starting point, one option might be to include this as an optional item in survey question banks so that it may be used by trusts with a particular interest in this area.

#### 4.4.4. “Overall, how would you rate the outcome of...” – summary of tested versions

##### Round one – standard wording and layout

**Q4.** Overall, how would you rate the outcome of your <experience: eg ‘visit to hospital’>?

- 1  Excellent
- 2  Very good
- 3  Good
- 4  Fair
- 5  Poor

##### Round two – amended question stem and response options for consistency with ‘overall, how would you describe your experience’ item

**Q4.** Overall, how would you describe the outcome of your <experience: eg ‘visit to hospital’>?

- 1  Very good
- 2  Good
- 3  Average
- 4  Poor
- 5  Very poor

##### Round three – merged into combined outcomes/experiences matrix

**Q3.** Thinking about <the episode>, how would you describe your....

	Very poor	Poor	Average	Good	Very good
<b>Outcome</b>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
<b>Experience</b>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>

## 4.5. Combined outcome/experience question

For the third round of testing we combined the two previous items on descriptions of overall outcome and experience into a single matrix layout asking people to evaluate these items together. As the main issues raised related to the layout of the items, it is appropriate to consider the testing of this separately to the earlier sections on outcomes and experiences<sup>1</sup>. The question tested was as follows:

**Q3.** Thinking about <the episode>, how would you describe your....

	Very poor	Poor	Average	Good	Very good
<b>Outcome</b>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
<b>Experience</b>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>

The aim of this layout was to encourage a clearer understanding of the differences between 'outcome' and 'experiences'. This was prompted by an observation in previous rounds that the two questions were better understood when adjacent to each other: moreover, a number of respondents spontaneously related their answers to the two questions when describing their choices to researchers, suggesting that it may be beneficial to explicitly group these items.

The combined layout worked well for many respondents. The majority of people interviewed gave different answers regarding their experiences and outcomes or, if ticking the same option for both, gave clear and distinct explanations of their reasoning for each case. For example, one interviewee told us that the difference between 'experience' and 'outcome' was that they reflected the short term (what it was like to be a patient and to be treated) versus the longer term (whether their health improved as a result of their treatment).

Unfortunately, there was also a significant minority of interviewees that struggled with the matrix layout. Five respondents (three in the GP setting and two in the inpatient setting) ticked only in the top row before moving on without answering the second part of the question. Although the relatively small number of interviews involved mean that we cannot generalise beyond this testing, it did appear that the problem was particularly pronounced amongst people for whom English was not their first language, and those with lower literacy. This would be in keeping with evidence that the greater complexity of matrix style questions can make them more problematic for respondents<sup>xv</sup>.

### 4.5.1. Conclusions

The matrix layout worked well for many respondents and often seemed to encourage clear and distinct interpretations of the two key terms (outcome and experience). However, the large proportion of people who only answered the first part of the question cannot be ignored and suggests that such a layout would lead to compromised data quality for the second part of the question. As such, we are unable to recommend this approach for surveys in the national programme. However, it should be noted that these surveys do not include other matrix questions, making this item more unusual in the tested questionnaires. It may be that the question would perform better in instruments that rely more on matrices, and there may be value in further testing this approach in such settings.

---

<sup>1</sup> Please note that issues raised in round three relating specifically to either of the terms "outcome" or "experience" are covered earlier in the sections relating to those questions.

## 4.6. Best and worst care received

### 4.6.1. Background

This was a new question – or rather pair of questions – designed for the specific purpose of being tested as a possible overarching indicator. The initial question involved two parts, one asking people to rate the best and the other the worst part of the care that they received.

**Q5a.** How would you rate the **best** care you received during you  
<experience: eg ‘visit to hospital’>?

- 1  Very good
- 2  Good
- 3  Average
- 4  Poor
- 5  Very poor
- 6  Don't know / can't say

**Q5b.** How would you rate the **worst** care you received during you  
<experience: eg ‘visit to hospital’>?

- 1  Very good
- 2  Good
- 3  Average
- 4  Poor
- 5  Very poor
- 6  Don't know / can't say

The rationale for this was to avoid a fundamental problem with summary measures. Single items can be problematic if respondents have had a very mixed experience, and they can be unsure how to answer if, say, a generally excellent experience was marred by one rude clinician. The idea of this pair of questions was to provide an outlet for respondents to indicate that their experience was not always of a single level of quality: for example, parts may have been very good whilst others were simply average. From an analytic perspective, this would, if successful, also provide an interesting new approach to understanding experience: we could still average responses to summarise experience, but we could also look at the differences between the two responses to understand when experience was consistent or mixed.

Although this indicator could potentially provide useful information and a more satisfying way of responding for participants, we were also conscious that it was somewhat more complex than most questions. Like the early “did you get the care that mattered to you” question it essentially involves two stages of consideration to answer: firstly respondents must think back and decide what the *best* or *worst* part of their care was and then secondly they must evaluate that against the scale provided. An additional issue is that we were unsure how respondents would react to

being asked to rate the worst care they received. We sought to address these issues in the testing.

#### 4.6.2. Cognitive testing

Overall, we tested three items following this 'range-reporting' approach. After the initial item tested poorly we substantially revised it to focus on whether people's expectations of care were met: two versions of this question were then tested in rounds two and three. Because the issues raised were different across the items, we describe findings from the first round and the latter two rounds separately.

##### Round one – issues raised

This pair of items polarised respondents. Around half of the interviewees answered the questions without difficulty and gave answers that reflected the design of the question. In several cases these respondents reflected on how they liked the approach being taken, and in some cases the questions seemed to better reflect people's experiences than any of the other overarching questions:

- **IP** One interviewee gave very positive answers to most other overarching questions but ticked "very poor" on the question about "worst care": this was the only response the person gave in this section that reflected the fact that she had initially been given an extremely distressing misdiagnosis.

Others, though, struggled greatly with the questions, particularly the latter question about "worst" aspects of care.

- **GP** Several interviewees indicated that they did not understand the concept of rating their "worst care" or otherwise how to differentiate between best and worst, particularly in the context of short visits to the GP.
- **IP** Several interviewees noted that all of their care was good so there was no "worst" point. Most eventually gave valid and consistent answers, but only after extended consideration.
- **MAT** Most interviewees struggled with the 'worst' question.
- **IP** One interviewee noted that it seemed incongruous to ask for a single evaluation about the best and worst aspects of care, as this seemed to imply that they occurred in equal proportion: this person wanted to indicate that almost all of their care was excellent, with a single very poor incident, but felt that ticking "very good" and "very poor" would give a different impression.

This latter point is of particular note, as this describes exactly the kind of experience that the questions were designed to be suited to: a mixed experience that included both good and bad points. That at least one interviewee continued to feel uncomfortable responding in this way (because of valid concerns about not wanting the frequency of good and bad experiences to be assumed to be equal) suggests that, overall, the questions did not work as intended.

##### Rounds two and three – issues raised

As noted above, this question was wholly revised for rounds two and three. Because of the concerns about frequency expressed in round one, we sought to develop an approach that would allow people to indicate how frequently care was very good or very poor. To ask this alongside the questions from the first round would be unrealistic because of the space required – effectively we would need four nested questions asking people to rate and describe the frequency of the best and worst aspects of care. Instead we took the approach of simply asking how often care deviated from some measure of what might be considered 'average': in this case, people's own expectations of care:

**Q5.** Thinking about all of the care and treatment you received, how often did this...

*Please tick once in each column*

	Not meet your expectations	Exceed your expectations
Always	1 <input type="checkbox"/>	1 <input type="checkbox"/>
Most of the time	2 <input type="checkbox"/>	2 <input type="checkbox"/>
Sometimes	3 <input type="checkbox"/>	3 <input type="checkbox"/>
Never	4 <input type="checkbox"/>	4 <input type="checkbox"/>

In the second round of testing this item again divided respondents. Around half struggled with it or did not understand it, whilst others indicated that they particularly liked it: one person told us that it was an “excellent question” that “made [me] think about all of [my] experience”. Where people struggled with the question, the most common issue was related to data quality, with people only answering half of the question. Many people placed a tick in only one column whilst only around half ticked in both columns as intended.

- **MH** One interviewee said that they did not like the question as it “made you think too much” and that more intuitive answers would be preferable.
- **IP** One interviewee, who eventually answered both columns as intended, initially assumed the question to be a trick – eg using scale inversion to identify inconsistent responses.
- **IP** One person refused the questions, asking “how can you answer two questions in one question?” (although they noted that if the questions were about two specific aspects of care this would have been fine).

Given that the majority of problems were related to missing responses, we considered that the layout of the question, with responses in columns, was problematic. For round three we transposed the question as follows:

**Q5.** Thinking about all of the care and treatment you received, how often did this...

*Please tick once in each row*

	Never	Some-times	Most of the time	Always
<b>Exceed your expectations</b>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>
<b>Fall short of your expectations</b>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>

Unfortunately, this did not successfully address the problems found in the previous round. Once again, the most common issue with the item was people providing an answer to only half of the question – for example ticking once to represent how often experiences met expectations but not indicating how often their experiences were not met. Only around half of the respondents placed a tick in each row as intended, and several appeared confused:



- **MH** One person noted that the question felt confusing because the order of the scales seemed to swap: a tick in the leftmost box was positive on one row and negative on the other. This person did, however, understand the question well and answer without difficulty.
- **IP** One interviewee struggled to answer the question properly, ticking 'never' in both rows to try to indicate that they had had a very bad experience.
- **GP** One interviewee ticked every box on the first row, but none on the second, attempting to indicate that their care was good.

In both rounds two and three there were also problems with the concept of expectations, with several people struggling to understand how to respond if care had simply met their expectations at all times: some of these interviewees said that they would have liked to see a 'just right' or 'met my expectations' column, but this would significantly increase the complexity of the question. Other interviewees simply objected to the term:

- **IP** One interviewee disliked the word 'expectations', saying that when going to hospital "the aim is just to get better". Oddly this person did not see 'getting better' as an 'expectation' *per se* – perhaps more as an objective.
- **GP** One interviewee answered the question as intended but said that it was "odd" and that they "hated it" because "expectations don't matter".
- **MAT** One interviewee wondered whether the question would be appropriate for people who might have unrealistic expectations: eg if someone had written a birth plan for themselves that was not medically appropriate.

### 4.6.3. Conclusions

Unfortunately, this approach did not work well. Although around half of respondents understood and answered well, and a few particularly favoured the approach, far too many other encountered serious problems when trying to answer the questions. It appears that the matrix style layout did not work at all well for presenting matched but opposite questions, whilst the approach of using separate but clearly related questions was also problematic. We do not recommend these items for use.

## 4.6.4. Best and worst care – summary of tested versions

### Round one – initial version

**Q5a.** How would you rate the **best** care you received during your  
<experience: eg 'visit to hospital'>?

- 1  Very good
- 2  Good
- 3  Average
- 4  Poor
- 5  Very poor
- 6  Don't know / can't say

**Q5b.** How would you rate the **worst** care you received during your  
<experience: eg 'visit to hospital'>?

- 1  Very good
- 2  Good
- 3  Average
- 4  Poor
- 5  Very poor
- 6  Don't know / can't say

### Round two – completely revised to focus on expectations being met

**Q5.** Thinking about all of the care and treatment you received, how often did this...

*Please tick once in each column*

	Not meet your expectations	Exceed your expectations
Always	1 <input type="checkbox"/>	1 <input type="checkbox"/>
Most of the time	2 <input type="checkbox"/>	2 <input type="checkbox"/>
Sometimes	3 <input type="checkbox"/>	3 <input type="checkbox"/>
Never	4 <input type="checkbox"/>	4 <input type="checkbox"/>

## Round three – response matrix transposed into rows instead of columns

**Q5.** Thinking about all of the care and treatment you received, how often did this...

*Please tick once in each row*

	Never	Some- times	Most of the time	Always
<b>Exceed</b> your expectations	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>
<b>Fall short</b> of your expectations	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>

## 4.7. “Overall...”

### 4.7.1. Background

As with the previous question on best and worst aspects of care, this was a new item designed for this project as a potential overarching measure of experience. Whilst the previous question consciously took the approach of adding complexity to seek to obtain more detailed feedback, this item by contrast was designed in a reductionist fashion to be as simple as possible. With the aim of ensuring the comparability of the item between different settings in mind, we used a one-word question stem to minimise the extent to which respondents would be encouraged to think about particular parts of their treatment. We also used a minimum set of response options to encourage consistency in interpretation of responses. The resulting question was minimal in design as follows:

**Q6.** Overall...

- 1  I would recommend <*this service*>
- 2  I would not recommend it
- 3  I'm not sure if I would recommend it

Whilst the question was designed to avoid leading people to consider certain aspects of care when responding, we were conscious that the lack of specificity might also cause problems – for example, if people were unsure of what the question was asking. Another concern was whether the response scale, with only one each of positive, negative, and ‘no opinion’ options, would be sensitive enough to allow people to answer in the ways in which they wanted. We sought to address these issues through the testing.

### 4.7.2. Cognitive testing

We tested three different versions of this item. Whilst the “Overall...” question stem remained unchanged throughout, we made significant changes to the response options between each round. These changes were designed to address the main issues that came up through the testing, which were:

- Response options
- Problems with the term ‘recommend’
- Other issues

### Response options

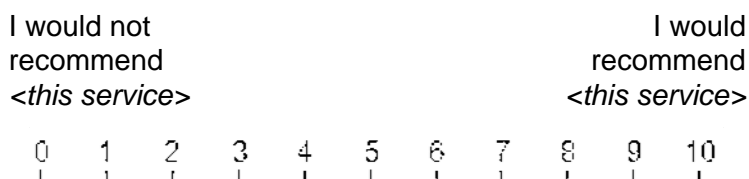
In the first round of testing we found the main difficulty to be related to the three response options. Almost all of the interviewees selected the “I would recommend <*this service*>” option, regardless of their answers to other questions. This seemed to reflect a reluctance to give the most negative option (“I would not recommend it”) even when they had had a bad experience. Indeed:

- **MAT** One interviewee ticked “I’m not sure if I would recommend it” because she felt that although she would not recommend the hospital she did not want to choose the most negative option.

In the case above the response options provided clearly failed to capture the person's experience, even though one of the unused response options closely matched their stated opinion. As responses of "I'm not sure if I would recommend it" would typically not be included in reporting of results, we expect that use of the question in this form would result in overwhelmingly positive responses. In the same round of testing, though, we found that several people told us that they preferred the longer eleven point scale used in the Net Promoter Score as it allowed a more precise response. We therefore added this scale to the question for round two, with the question appearing as follows:

**Q6.** Overall...

*Please circle the appropriate number*



This response scale seemed more useful, but we encountered a new issue in round two, where we noted reluctance from some interviewees to select the most positive option on the scale. A number of people indicated that they would not be prepared to score a '10' for recommendation under any circumstances:

- **MAT** One woman gave a score of '8', which she described as a good score for her as she would "never" give a 10.
- **MH** Two people told us that they would be unlikely or unwilling to give a score of 10: one said that "nothing is 100%", whilst another indicated that they "would never give anyone a 10" because this would imply an unachievable level of perfection.
- **IP** One interviewee told us that "I never give 10s [because] it's bad luck". Another gave a score of 9 because they said there was "always room for improvement">
- **GP** One interviewee selected 9, saying that they would not select 10 as they would not want their practice to become oversubscribed (eg because it was too emphatically endorsed and therefore became excessively popular).

For the third round of testing, we replaced the term 'recommend' with 'experience' (see below). We did not find the same problems with the scale as in round two, and indeed the full range seemed to be well used. One remaining issue did recur, though, in that two respondents gave non-integer responses (an inpatient selecting 8.5 and a GP patient selecting 9.5). Both of these people went on to give integer responses for the subsequent NPS question, where the line beneath the response options had been removed – suggesting that it might be preferable to similarly remove the line from beneath the options on this item.

### Problems with the term 'recommend'

As with the Net Promoter Score, we found that many interviewees objected to the use of the term 'recommend' in this item. Problems occurred consistently across all questionnaires and many of the objections raised were similar to those for the Net Promoter Score. However, it should be noted that concerns were expressed regardless of whether this item or the NPS appeared first:

- **IP** Two interviewees refused both this item and the Net Promoter Score, giving the same reasons for both items. One said that they "didn't go around recommending hospitals" whilst the other said that they "wouldn't recommend hurting your hip" and that the concept of 'recommendation' was inappropriate for emergency admissions.

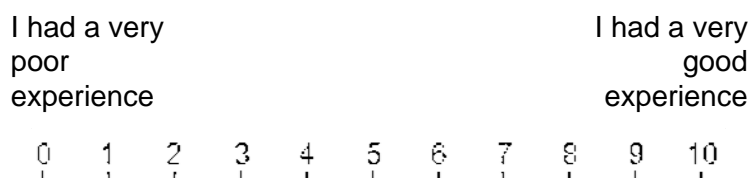
- **MH** One interviewee refused to answer the question, saying that they were “not in the business of recommending mental health services”.
- **GP** One interviewee told us that as they lived in a rural village there was no real choice of GP practice in their area: consequently they felt the word ‘recommend’ was inappropriate.
- **MAT** Two women found the concept of recommending services strange. One volunteered that “you would recommend a mobile phone” but not medical care, whilst another indicated that as services were neither paid for nor “optional” that it was strange to consider recommending them.
- **GP** In the second round of testing one interviewee circled ‘9’, saying that they had a very good experience overall. However, they noted that they disliked the question because one “can’t quantify recommendation”.

As well as objections to the term ‘recommend’ we also found that a minority of people misattributed the question to be about something other than the services they had used. This was similar to the problem identified in testing of the Net Promoter Score, although it seemed somewhat less prevalent with this item:

- **MH** One interviewee understood the question to be about their recommendation for treatment, rather than the specific services they had used. They gave a positive response, explaining that they would definitely recommend that people with similar problems seek help.
- **MH** One interviewee said that they were answering about the NHS as a whole rather than the services that they had used.

Because of these objections and attributions, we revised the item for round three to look at ‘experiences’ instead of ‘recommendations’:

**Q6.** Overall...



## Other issues

The question tested very well in the third round of testing. Only two people raised any objection to the use of the term “experience”:

- **MH** One interviewee stated that having a mental health condition was clearly not a “good experience” in and of itself. Nonetheless, they understood that the question was asking about the services they received and ticked 10, which was consistent with their responses elsewhere and what they described as a very positive experience of care.
- **GP** One interviewee read the response scale and asked “experience of what?”. This person was, however, still able to answer the question appropriately, and it should be noted that they had difficulties with written English and struggled with many items in the questionnaire.

Beyond these minor points, the question seemed to work well and there was good evidence in rounds one and two that people were treating the item as intended: that is, an overarching measure asking for a summation of everything about their experience of care. For example:



#### 4.7.4. “Overall...” – summary of tested versions

##### Round one – initial version

Q6. Overall...

- 1  I would recommend <this service>
- 2  I would not recommend it
- 3  I'm not sure if I would recommend it

##### Round two – response options revised to 11 point scale

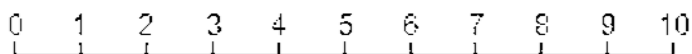
Q6. Overall...

*Please circle the appropriate number*

I would not  
recommend  
<this service>

I would  
recommend  
<this service>

0 1 2 3 4 5 6 7 8 9 10



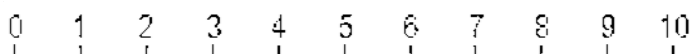
##### Round three – ‘recommend’ replaced with ‘experience’

Q6. Overall...

I had a very  
poor  
experience

I had a very  
good  
experience

0 1 2 3 4 5 6 7 8 9 10





## 5 Discussion

This was a large scale project with a considerable number of interviews undertaken with people from a wide range of backgrounds. Because of the numbers of interviews involved we are confident that the testing provides a robust assessment of the qualities of the items investigated. Indeed, the numbers within each sector are themselves large and give a good impression of the appropriateness of the questions for different care sectors.

Six items were tested over the course of the project, each undergoing at least some revision following the first and second rounds of interviews. A number of interviewers were involved in each round of testing, and their contributions at end-of-round debriefings were invaluable for understanding issues arising from the testing and appropriately refining questions.

Through the testing, it became apparent that some of the items were inappropriate or did not work well. These included:

- The Net Promoter Score. This caused problems in all rounds of testing because interviewees objected to and misunderstood the term 'recommend'. As this is a fundamental part of the question it was not possible to address this through revisions to the wording, and we conclude that the NPS is unsuitable for use in NHS settings.
- The new 'best and worst aspects of care' question. The two part design confused respondents despite changing the layout, and some people objected to being explicitly asked about parts of care that were poor or fell short of their expectations. Although the question worked very well for some other interviewees, we conclude that it is not suitable for use in the national programme.
- A question on whether people "got the care that mattered to [them]". Although most interviewees answered the item with ease we found that there was great variance in how the question was interpreted and evaluated. The wide range of issues being considered by different people when responding to the item limits its value for service improvement or performance assessment. In particular, only a minority of people responded in a summative fashion, with many others responding based on a single specific aspect of care. This means it would be difficult if not impossible to interpret whether a poor score on this question reflected atypically bad experiences overall or a single weak point in the provision of care. Consequently we do not recommend this item as an overarching measure of patient experience.

Other items performed better in terms of being clearly and consistently understood, but

- A question on overall descriptions of experiences worked reasonably well, but there were issues relating to the use of the word 'experience' in the question stem (for example, because needing hospital treatment is "never a good experience") and the sufficiency of the response options. In particular, we were concerned at likely floor effects in use of the negative response options, which could create exaggeratedly positive reports of care.
- A similar question on overall outcomes worked well, particularly when contrasted to a question on experiences. This item was evaluated based on clinical outcomes, which means that it is not appropriate as an overarching measure of patient experience. However, we consider this a strong and valuable question, and it could be included in future surveys to provide useful supplementary information and to allow comparison of people's experiences and outcomes.

Ultimately, we found that the question that worked best across the range of sectors was a simple 'overall' question with an 11 point Likert scale:

- The "Overall..." question tested in round three was well understood by interviewees and proved successful as a measure of overall experience. More than any other item

this seemed to encourage people to think about their whole experience and to provide a summative assessment of how positive this was. Moreover, the minimal wording of the question stem meant that this item adapts easily across settings and was well understood by respondents. We recommend this item for use in future surveys.

Given that interviews have been undertaken in such a diverse range of settings, and the recommended question found to be suitable in all of these, we expect that it will work similarly well in the majority of other NHS settings. The rationale for this is that we consider the settings chosen to represent extremes of care in terms of their differences from each other: and because we cover a mixture of elective and emergency, physical and mental, 'condition' specific and general, and acute and primary care, we feel that most untested settings bear at least some similarity to part of the work conducted here. In spite of this, caution should be taken in transferring the recommended item – or indeed any item – to very unusual or unique care settings.

Although this work involved a very large number of cognitive interviews and we are confident of the robustness of the findings relating to the content validity of the tested items, the project is not without its limitations. Because each item was tested up to a maximum of 26 times before revision we did not collect quantitative data from the project. This means that we cannot test the statistical properties of the questions at this time. Should the recommended question be used in national surveys or a pilot survey then a priority for future research should be to investigate these statistical properties to confirm that it provides a suitable measure for discriminating between providers.

Another limitation related to the lack of quantitative data is that we cannot statistically compare results between survey settings. This means that whilst we are confident that this work demonstrates the acceptability of the recommended item across a range of settings, we are unable to test the statistical comparability across settings. A particular challenge in doing this is that different kinds of care providers will have very different case mixes of patients. We already know from work on the national survey programme that different demographic groups tend to respond differently to surveys, so, at a minimum, care will need to be taken to balance for differences that are attributable purely to demographic groups. There is also the further risk that unmeasured characteristics correlated to particular settings may influence response, and this should be acknowledged in future work.

Finally, it is important to note that we are not recommending that the "Overall..." question tested here be used in isolation. Generally speaking, we would not advocate or recommend single-item questionnaires as these will almost always be highly inefficient and of limited value, particularly from a service improvement perspective. The most effective way of using this overarching measure will almost always be alongside other items covering specific reportable experiences: this will allow users to 'drill down' into the data and form a clearer understanding of the issues causing people to have good or poor experiences.

With these caveats in mind, we are nonetheless confident that the recommended "Overall..." question constitutes a sound and content valid way of measuring patients experiences of care in an overarching sense and across a broad range of settings. This should be added to future surveys, and further statistical analysis should be undertaken once data has been gathered to measure the statistical properties of the question.

## 6 Appendices: interim reports

### 6.1. Cognitive testing interim report: round one

This report covers the first of three rounds of cognitive testing of overarching questions. This testing took place between the 17<sup>th</sup> and 28<sup>th</sup> October 2011, and included a total of 29 interviews across four topic areas. In this document we briefly describe the range of interviews completed and questions used before looking in detail at findings around each of the six test items. Suggested revisions for round two are summarised in the enclosed appendix.

**Acknowledgment:** interviews relating to maternity services were conducted by colleagues at CQC. We are grateful for their involvement and input, and this document draws heavily on their comments when discussing findings from the maternity questionnaire.

#### 6.1.1. Interviews completed

In total, 29 interviews were conducted as part of round one. A summary of interviews is presented in table one, below.

Topic area/ questionnaire	Number of interviews completed	Notes
GP services	10	Ten interviews were completed rather than eight as planned because of unintentional over-recruitment by the agency responsible for identifying interviewees.
Acute inpatients	8	
Community mental health	6	See note below.
Maternity services	5	Interviews conducted by staff at CQC.
<b>All topic areas</b>	<b>29</b>	

### Community mental health interviews

Recruitment of interviewees for the community mental health questionnaire was difficult because it was not possible to complete interviews with a sizable minority of all those who originally agreed to take part. Only six interviews were fully completed, but in addition to this:

One interview was terminated at an early stage because the interviewee was distressed and unable to answer questions.

One interviewee failed to attend on two occasions.

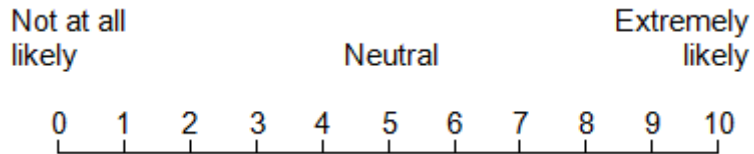
Two interviewees cancelled appointments at short notice.

Unfortunately because of these incidents it was not possible to recruit additional interviewees within the required timeframe. Nonetheless, we are confident that sufficient coverage of the community mental health setting has been achieved for round one.

## 6.1.2. Questions tested

The following questions were tested in round one:

**Q1.** How likely is it that you would recommend *<this service>* to a friend or colleague?



**Q2.** On reflection, did you get the care that mattered to you?

- 1  At all times
- 2  Most of the time
- 3  Sometimes
- 4  Rarely
- 5  Never
- 6  Do not wish to answer

**Q3.** Overall, how would you describe your experience?

- 1  Very good
- 2  Fairly good
- 3  Neither good nor poor
- 4  Fairly poor
- 5  Very poor

**Q4.** Overall, how would you rate the outcome of your  
<experience: eg 'visit to hospital'>?

- 1  Excellent
- 2  Very good
- 3  Good
- 4  Fair
- 5  Poor

**NB: please note that question five is considered a two part question.**

**Q5a.** How would you rate the best care you received during  
your <experience: eg 'visit to hospital'>?

- 1  Very good
- 2  Good
- 3  Average
- 4  Poor
- 5  Very poor
- 6  Don't know / can't say

**Q5b.** How would you rate the **worst** care you received during  
your <experience: eg 'visit to hospital'>?

- 1  Very good
- 2  Good
- 3  Average
- 4  Poor
- 5  Very poor
- 6  Don't know / can't say

**Q6.** Overall...

- 1  I would recommend <this service>
- 2  I would not recommend it
- 3  I'm not sure if I would recommend it

### 6.1.3. Findings

A large number of interviews were undertaken across a range of services. Participants came from a wide range of backgrounds and had very different stories to tell in terms of their experiences. For convenience, this report summarises findings by question.

#### Q1. Net promoter score

This item tested fairly well overall. Many interviewees noted that they liked the greater number of options afforded by a longer scale, although it was not clear that this scale was evaluated consistently by different people. There was some variation across topic areas, too: the item was answered more readily by people who had used GP or maternity services, but a number of inpatients and mental health service users felt uncomfortable with the concept of recommending health services.

- **IP** Two interviewees refused to answer the question because they felt it was inappropriate. One stated that “I don’t go around recommending hospitals” whilst another told us that they had an emergency admission and “wouldn’t recommend hurting your hip”: this person also noted that the idea of ‘recommendation’ implies that people have a choice about the services they use, which is not the case for emergency admissions.
- **MH** One interviewee immediately circled the option 10 – but on probing stated that this was not a reaction to the quality of care that they had received but rather reflected the fact that they would encourage people with mental health issues to seek professional help. In other words, they had answered in relation to the (group of) profession(s) as a whole, rather than about their own care providers.
- **MH** One interviewee felt that their response to this question would be related to their condition and would therefore be more indicative of their mental health rather than the quality of services.

Whilst many interviewees liked the flexibility of a longer scale with more options, others were put off by this or experienced difficulties in responding as intended:

- **GP** One interviewee stated a dislike for the scale, saying they preferred written (categorical) questions.
- **GP** One respondent multicoded this item, circling options 5, 6, and 8, as well as the word ‘neutral’. They explained that the ‘8’ referred to experiences in hospital, whilst their other selections were regarding their GP practice.
- **GP** Two further respondents stated that they would like to choose ‘between’ numbers (ie select non-integer values, such as 4.5). They were nonetheless able to give valid responses without prompting.
- **IP** One interviewee said that they wanted to select ‘8.5’ as an answer, rather than 8 or 9. They eventually gave a valid answer.
- **IP** One interviewee answered in relation to all of their hospital experiences, not just their recent one: they had attended the same hospital on numerous occasions and thought it odd to be asked to recommend it based purely on a single episode.

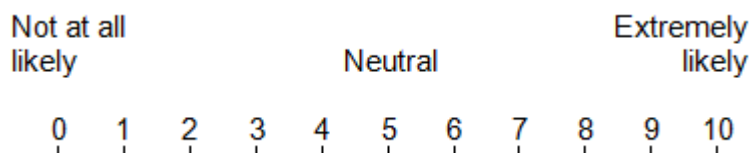
Generally, most interviewees seemed to consider a number of factors when responding to this question, but a minority answered specifically in relation to clinical factors such as the outcome of their treatment.

- **IP** One interviewee selected ‘8’ despite describing what was clearly a very upsetting hospital experience (including misdiagnosis and lack of information). They stated that this was because they now felt physically better.

## Revisions for round two

On the whole, it appeared that this item was working well in the GP and maternity settings, but poorly in settings where choice is less apparent – ie inpatient care and mental health. We therefore recommend revising this item as follows:

**Q1.** How likely is it that you would recommend <this service> to a friend or colleague if they needed similar care or treatment?



Our intention is that this addition should help to contextualise the question, and to make it clear to respondents that we are interested in their recommendation of the care that they received. This should also address the issue of people feeling that the item is inappropriate, as the revised words add the presupposition that the care is required.

## Q2. “On reflection, did you get the care that mattered to you?”

Interviewees gave very mixed reactions to this item. Many were able to respond quickly and gave clear, reasoned explanations of how they had arrived at their answer. Some answered with regard to numerous aspects of care, whilst others picked the one factor that mattered most to them.

Variously, interviewees said that what mattered to them was:

- **GP** Having enough time with a doctor
- **GP** The clinical competence of their GP
- **GP** Their overall care
- **IP** Getting prompt attention
- **IP** Clinical outcomes or “getting better”
- **IP** “Everything mattered”
- **MH** Rapid attention and practical interventions
- **MH** Expectations being met
- **MAT** Being well looked after
- **MAT** Having all of the necessary tests and checks
- **MAT** Being given good information
- **MAT** Caring staff

Some interviewees were very clear about what they saw the question as *not* being about, or what they thought they were meant to exclude:

- **GP** One interviewee said that they were confused as they weren’t sure whether they could answer based on having had problems getting through to their practice on the phone.
- **GP** One interviewee assumed that the question was about the GP rather than the GP practice, and therefore disregarded experiences with other members of staff.

- **GP** One interviewee told us that they excluded factors like politeness of staff and cleanliness of the practice as they did not consider these to be parts of “care”.
- **IP** Two interviewees answered based only on clinical outcomes.
- **IP** By contrast, one said that “care relates more to people [ie relationships with staff] than to treatment”.

The question seemed most problematic in the mental health setting, where people tended to be in receipt of long-term care from a variety of sources:

- **MH** One interviewee was unclear about whether to include care received from their GP, which they thought was particularly important to them.
- **MH** One interviewee indicated that what mattered to them varied over time, whereas the item seemed to be asking about a particular episode.

Several interviewees expressed surprise at the “do not wish to answer” option for this question, stating that it was odd that this was included for this item but not for others. None of the people we interviewed selected this item, so we recommend removing it.

At least one interviewee completely misunderstood the question:

- **GP** One interviewee said that the question was “fuzzy” and that they took it to be an indirect way of asking if they “had ever been upset”

## Revisions for round two

Overall, this item caused relatively few problems for interviewees in the maternity, inpatient, and GP settings – although it was slightly more difficult for people in the mental health group who received longer term care. Whilst the majority of people were able to respond without difficulty, though, it appeared that there might be some significant differences in the interpretation of the question – particularly with what people considered to be part of “care”.

Because of this, and because it was clear that what mattered to different people varied considerably (limiting the value of this question from a service improvement perspective), our concerns about the item are more fundamental. We therefore recommend retaining it for round two with only a small change – dropping the “do not wish to answer” response option – so that we can build a clearer understanding of this. In the second round of interviews, we will also specifically probe on what does and does not constitute part of care from the perspective of interviewees.

**Q2.** On reflection, did you get the care that mattered to you?

1  At all times

2  Most of the time

3  Sometimes

4  Rarely

5  Never

6  ~~Do not wish to answer~~



### Q3. “Overall, how would you describe your experience”.

We encountered few problems when testing this item: it generally appeared to work well across the range of settings. For the most part, the term “experience” was clearly understood, and interpreted consistently with its intention. For example:

- **IP** One interviewee described the question as asking about “more than just [*the hospital*] doing its job” in a medical sense: another echoed this by saying that they considered everything that happened in the hospital when evaluating their response.

There were some issues in understanding, however:

- **GP** One interviewee was slightly confused as to whether the question was asking about the experience with the GP or the experience of the practice.
- **MAT** Perhaps more significantly, one interviewee said that it was not clear that this question referred to ‘care’ as it did not specify “what experience” was being asked about.
- **MAT** Another recent mother indicated that she felt there were two interpretations – her experience in the context of her life (ie childbirth and labour being one of the most unpleasant experiences she had had) and then the experience of how she was cared for by the service.

Whilst the question itself typically worked well, one of the response options was extremely problematic. A sizable minority of all interviewees (roughly one in three) said that they interpreted “neither good nor poor” to mean “I have no opinion”, whilst the rest felt that this meant “neutral” or “average”. These interpretations are clearly quite different, creating problems for how the data should be used.

### Revisions for round two

Generally, this question worked well. In order to try and avoid the problems identified around how to contextualise ‘experience’, though, we will add introductory text to the question as below: we will also take the opportunity in round two to probe on how people understand the term “experience”. We will also replace the third response option with “average” to make clear that this indicates neutral rather than no opinion.

***Still thinking about <experience: eg “the NHS care you received during your labour and the birth”>***

**Q3.** Overall, how would you describe your experience?

- 1  Very good
- 2  Fairly good
- 3  ~~Neither good nor poor~~ Average
- 4  Fairly poor
- 5  Very poor

## Q4. “Overall, how would you rate the outcome of...”.

The vast majority of interviewees interpreted “outcome” to refer to the clinical effectiveness of their care or treatment or, in other words, whether they had ‘gotten better’.

The item seemed particularly suited to the inpatient setting, with respondents consistently interpreting the question as being about clinical outcomes and answering appropriately: moreover, it was clear that interviewees were able to separate out the clinical outcome of their care from other aspects of their experience.

- **IP** One interviewee selected “excellent” without hesitation and immediately explained that this was because “I’m just fine now”. This person noted several serious issues with the care that they received (initially being misdiagnosed, and later being given no information about the outcome of a surgical procedure for two months) but said that they did not think this was relevant to the question.
- **IP** One interviewee initially ticked “fair” but then amended this to “excellent” after noticing the word “outcome” in the question. They explained that their overall stay in hospital was ‘average’ but that they were “a lot better” as a result of it.

Whilst the item was very well suited to the inpatient setting, the term outcome caused difficulties in the other surveys:

- **GP** Three interviews expressed confusion at the term ‘outcome’ as they considered their care to be ongoing. One noted that they considered the outcome of their most recent appointment to be a referral to a hospital consultant, and they were not sure how this should be evaluated.
- **MAT** Although this question was carefully worded to try and avoid such issues, one interviewee expressed amusement at the word “outcome” as she said that her “baby was the outcome”.
- **MH** As with the GP setting, some interviewees were unclear about responding to this when care was ongoing. One indicated that they interpreted “outcome” as asking “how I’m doing now”.

A number of interviewees raised concerns relating to the response scale in use.

- **GP** One interviewee noted that the scale was unbalanced, with three positive and two negative options: they felt that this was inappropriate.
- **IP** One interviewee noted that the scale was unbalanced. Another stated that they thought “excellent” required a “very high” standard.
- **MH** One interviewee felt unable to select the top box as they thought this implied “perfection”.

## Revisions for round two

The item tested well in the inpatient setting but less so in other questionnaires. It is not clear that this can be addressed, as the term ‘outcome’ is both fundamental to the nature of the question and the main cause of confusion for some respondents. It may therefore be that this question would be suitable as an addition to the inpatient survey to cover clinical outcomes (and to allow analysis of the relationship between this and more relational aspects of patient experience), albeit not as an overarching measure of ‘experience’.

Rather than drop this item, we feel that further testing would be beneficial. Firstly, we will replace the unbalanced scale with a balanced five-point scale. Secondly, we will slightly amend the question stem to replace the word “rate” with “describe” to determine whether this less evaluative phrasing influences people’s understanding of the question in relation to long-term and continuous care.

**Q4.** Overall, how would you ~~describe rate~~ the outcome of your <experience: eg 'visit to hospital'>?

1  ~~Excellent~~ Very good

2  ~~Very good~~ Good

3  ~~Good~~ Average

4  ~~Fair~~ Poor

5  ~~Poor~~ Very poor

## Q5. “How would you rate the best/worst care you received...”.

This pair of items divided respondents. Around half of the interviewees answered the questions without difficulty and seemed to give answers that reflected the design of the question. Others, though, struggled greatly with the questions, particularly the latter question about “worst” aspects of care.

- **GP** Several interviewees indicated that they did not understand the concept of rating their “worst care” or otherwise how to differentiate between best and worst, particularly in the context of short visits to the GP.
- **IP** Several interviewees noted that all of their care was good so there was no “worst” point. Most eventually gave valid and consistent answers, but only after extended consideration.
- **IP** One interviewee noted that it seemed incongruous to ask for a single evaluation about the best and worst aspects of care, as this seemed to imply that they occurred in equal proportion: this person wanted to indicate that almost all of their care was excellent, with a single very poor incident, but felt that ticking “very good” and “very poor” would give a different impression.
- **MAT** Most interviewees struggled with the ‘worst’ question.

However, there were some cases where the questions were answered well and seemed to cast a different perspective compared to the summative questions. For example:

- **IP** One interviewee gave very positive answers to most other overarching questions but ticked “very poor” on the question about “worst care”: this was the only response the person gave in this section that reflected the fact that she had initially been given an extremely distressing misdiagnosis.

## Revisions for round two

The questions tested did not work well, but did generate evidence that there may be merit in further exploring the idea of asking people to tell us about the range of their experience. We therefore feel that a complete revision of these questions would be appropriate, as follows:

### Q5. Thinking about all of the care and treatment you received, how often did this...

*Please tick once in each column*

	Not meet your expectations	Exceed your expectations
Always	1 <input type="checkbox"/>	1 <input type="checkbox"/>
Most of the time	2 <input type="checkbox"/>	2 <input type="checkbox"/>
Sometimes	3 <input type="checkbox"/>	3 <input type="checkbox"/>
Never	4 <input type="checkbox"/>	4 <input type="checkbox"/>

We recognise that the matrix design may prove complex for some readers but feel there is value in testing this rather than two separate questions.

## Q6. “Overall...”

In general we found that the question stem of this item was unproblematic and worked well across all settings. As with the net promoter score, though, a number of interviewees objected to the use of the word ‘recommend’:

- **IP** Two interviewees refused the question: these were the same people that refused the net promoter score question, and the same reasons were given.
- **MH** One interviewee refused the question as they said they were “not in the business of recommending mental health services”.

Aside from the people who refused to answer the question, there were no other significant problems with understanding or interpreting the question: interviewees consistently took this to be asking for an overall evaluation of the quality of the services that they had received.

However, whilst the question stem worked well there were significant problems with the response options. Almost all respondents ticked the “I would recommend...” option, regardless of how they answered other questions. This seemed to reflect a reluctance to give the most negative option (“I would not recommend it”) and a lack of intermediate options. For example:

- **MAT** One interviewee ticked “I’m not sure if I would recommend it” because she felt that although she would not recommend the hospital she did not want to choose the most negative option.

## Revisions for round two

Given that the question stem worked well but the scale seemed inappropriate, we recommend replacing the scale for this question with an 11-point scale as used in the net promoter score. This should balance the qualities of the two questions by incorporating the highly differentiated scale from the NPS with a less problematic question stem.

### Q6. Overall...

*Please circle the appropriate number*

I would not  
recommend  
<this service>

I would  
recommend  
<this service>

0 1 2 3 4 5 6 7 8 9 10

1-  I would recommend <this service>

2-  I would not recommend it

3-  I'm not sure if I would recommend it



**Q4.** Overall, how would you ~~describe rate~~ the outcome of your <experience: eg 'visit to hospital'>?

- 1  ~~Excellent~~ Very good
- 2  ~~Very good~~ Good
- 3  ~~Good~~ Average
- 4  ~~Fair~~ Poor
- 5  ~~Poor~~ Very poor

**NB:** please note that question five is considered a two part question.

**Q5.** Thinking about all of the care and treatment you received, how often did this...

*Please tick once in each column*

	Not meet your expectations	Exceed your expectations
Always	1 <input type="checkbox"/>	1 <input type="checkbox"/>
Most of the time	2 <input type="checkbox"/>	2 <input type="checkbox"/>
Sometimes	3 <input type="checkbox"/>	3 <input type="checkbox"/>
Never	4 <input type="checkbox"/>	4 <input type="checkbox"/>

~~Q5a. How would you rate the **best** care you received during your <experience: eg 'visit to hospital'>?~~

~~1-  Very good~~

~~2-  Good~~

~~3-  Average~~

~~4-  Poor~~

~~5-  Very poor~~

~~Q5b. How would you rate the **worst** care you received during your <experience: eg 'visit to hospital'>?~~

~~1-  Very good~~

~~2-  Good~~

~~3-  Average~~

~~4-  Poor~~

~~5-  Very poor~~

Q6. Overall...

*Please circle the appropriate number*

I would not  
recommend  
<this service>

I would  
recommend  
<this service>

0 1 2 3 4 5 6 7 8 9 10

~~1-  I would recommend <this service>~~

~~2-  I would not recommend it~~

~~3-  I'm not sure if I would recommend it~~



## 6.2. Cognitive testing interim report: round two

This report covers the second of three rounds of cognitive testing of overarching questions. This testing took place between the 7<sup>th</sup> and 18<sup>th</sup> November 2011, and included a total of 28 interviews across four topic areas. In this document we briefly describe the range of interviews completed and questions used before looking in detail at findings around each of the six test items. Suggested revisions for round three are summarised in the enclosed appendix.

**Acknowledgment:** interviews relating to maternity services were conducted by colleagues at CQC. We are grateful for their involvement and input, and this document draws heavily on their comments when discussing findings from the maternity questionnaire.

### 6.2.1. Interviews completed

In total, 28 interviews were conducted as part of round two.

A summary of interviews is presented in table one, below.

Topic area/ questionnaire	Number of interviews completed	Notes
GP services	8	
Acute inpatients	8	
Community mental health	7	See note below.
Maternity services	5	Interviews conducted by staff at CQC.
<b>All topic areas</b>	<b>28</b>	

### Community mental health interviews

As in the first round of testing we found recruitment of interviewees for the community mental health questionnaire to be problematic. There were a number of issues that led to only seven usable interviews being completed:

One interview was completed but disregarded because it became clear that the interviewee had not used any *community* mental health services. They had received treatment for psychosis whilst in prison but had received no further care following their release.

One interviewee failed to attend on two occasions.

Because of these difficulties, additional recruitment was undertaken following an alternative route: instead of using an agency, the Picker Institute advertised directly for potential candidates. This approach showed more promise and will be used as default for the mental health group in round three. Nonetheless, we are confident that sufficient coverage of the community mental health setting has been achieved for round two.



**Q4.** Overall, how would you rate the outcome of your <experience: eg 'visit to hospital'>?

- 1  Very good
- 2  Good
- 3  Average
- 4  Poor
- 5  Very poor

**NB: please note that question five is considered a two part question.**

**Q5.** Thinking about all of the care and treatment you received, how often did this...

*Please tick once in each column*

	Not meet your expectations	Exceed your expectations
Always	1 <input type="checkbox"/>	1 <input type="checkbox"/>
Most of the time	2 <input type="checkbox"/>	2 <input type="checkbox"/>
Sometimes	3 <input type="checkbox"/>	3 <input type="checkbox"/>
Never	4 <input type="checkbox"/>	4 <input type="checkbox"/>

**Q6.** Overall...

*Please circle the appropriate number*

I would not recommend  
<this service>

I would recommend  
<this service>

0 1 2 3 4 5 6 7 8 9 10

### 6.2.3. Findings

A large number of interviews were undertaken across a range of services. Participants came from a wide range of backgrounds and had very different stories to tell in terms of their experiences. For convenience, this report summarises findings by question.

#### Q1. Net promoter score

Following round one of testing, we had adjusted the wording of this question to include the suffix "...if they needed similar care or treatment". The purpose of this change was to address the issue that many respondents felt that the question was akin to 'recommending illness'.

This item tested fairly well overall in round two, although as in round one it was not clear that this scale was evaluated consistently by different people. As in round one, a number of people also gave unintended responses – eg selecting more than one number or writing in non-integer values:

- **MH** One interviewee said that they "would recommend" the services they received, but wrote in a score of "6.5" because "nothing's perfect".
- **MH** Another interviewee wrote in "7.5" because they preferred to "think in quarters".
- **GP** One interviewee wrote in "8.5".

In spite of the change to the question stem, a minority of interviewees still expressed their dislike of the focus on 'recommendations' – particularly regarding acute hospital care:

- **IP** One interviewee stated that "I don't recommend hospitals... it's ridiculous" – but nevertheless circled option 9.
- **MAT** One interviewee commented on the wording, saying that 'you would not recommend NHS care because you don't pay for it'.
- **IP** Another interviewee refused the question, stating that as they had not seen every doctor in the hospital it would be "stupid" to recommend the hospital based only on the doctors that they had seen.

Although this latter interviewee refused to answer the question, others raised similar concerns.

- **GP** One interviewee struggled with the question because they felt that they would be "more likely to recommend my GP than the practice". They also noted that they thought the word 'recommend' was inappropriate, as there was no real choice of GP practice in their area (a rural village). They eventually ticked option 5.
- **GP** An interviewee receiving regular treatment related to alcohol dependency stated that they would recommend their GP for people needing similar treatment, but not for those with other conditions. In spite of the suffix "...if they needed similar care or treatment" they hesitated over this question as they felt that it was looking for a more general recommendation.

There were some specific problems with this question in the maternity survey, which seemed to relate to the wording of the question stem. In particular, the use of the term 'pregnancy' in this the initial item about antenatal services may have led some women to think beyond the antenatal care received, with one interviewee noting that she was considering her experiences of emergency care when answering the question.

#### Revisions for round three

This item worked reasonably well for most interviewees, although a minority continued to express concerns with the use of the term 'recommend'. We consider it unlikely that these concerns could be addressed without fundamentally changing the question, so we do not propose any changes relating to this at this stage.

Another problem that reappeared from the first round was multicoding or giving invalid responses (eg 7.5). It seems that using a scale with more options encourages some respondents to want to give more precise answers and to assume that these will be usable. To tackle this, we recommend testing the question with the horizontal line removed from beneath the response numbers: this may visually deemphasize the scale.

**Q1.** How likely is it that you would recommend <this service> to a friend or colleague if they needed similar care or treatment?

*Please circle the appropriate number*

Not at all likely						Neutral						Extremely likely
0	1	2	3	4	5	6	7	8	9	10		

## Q2. “On reflection, did you get the care that mattered to you?”

As in the first round of testing, interviewees gave very mixed reactions to this item. Many were able to respond quickly and gave clear, reasoned explanations of how they had arrived at their answer. Some answered with regard to numerous aspects of care, whilst others picked the one factor that matter most to them.

Variously, interviewees said that what mattered to them was:

- **GP** Simply being “always cared for”, in spite of other problems
- **GP** The reassurance provided by their GP and the GP’s bedside manner.
- **GP** “Not being dissatisfied”
- **GP** Being treated with respect and not patronised.
- **1 GP & 3 IP** Clinical outcomes ; “getting well”
- **IP** Relational aspects of care; being treated well by staff.
- **IP** Having trust in the staff
- **2 MH** Having people to talk to, or being listened to
- **MH** Receiving personalised care
- **MH** Free access to services
- **MAT** Being told what was happening and getting answers to questions
- **MAT** Getting help when needed
- **MAT** Having frequent antenatal checks

In contrast to the first round of testing, we observed fewer problems with people assuming that only certain aspects of care should be included or excluded. This may be related to the ordering of the questionnaire. In round one this question was preceded by the net promoter score, which may have encouraged respondents to think about it in particular ways. Placed as the first question in the ‘overall’ section it seemed to perform better.

## Revisions for round three

We encountered no new problems with interviewees' ability to answer the question in round two, and therefore recommend no changes to the item itself for the third round. However, there was some concern that the question appeared to handle very differently when placed first in the 'overall' set rather than after the net promoter score. This raises the possibility that the question may be particularly susceptible to order effects, perhaps because it is so open-ended. For round three we will therefore move it to a later position in the set of overall questions.

More generally, and as noted in the previous round, it was clear that what mattered to different people varied considerably. This limits the value of this question from a service improvement perspective, and so our concerns about this item are more fundamental.

**Q2.** On reflection, did you get the care that mattered to you?

- 1  At all times
- 2  Most of the time
- 3  Sometimes
- 4  Rarely
- 5  Never

## Q3. "Overall, how would you describe your experience".

We encountered few problems when testing this item: it generally appeared to work well across the range of settings. As in round one, the term "experience" was typically clearly understood, and interpreted consistently with its intention – but one inpatient did say that they did not understand what the term was referring to. Some interviewees also expressed surprise at the question: two inpatients stated that a visit to hospital is never a good experience (but nevertheless ticked 'very good' and 'fairly good' respectively).

The question proved slightly more problematic in the maternity survey, particularly in the context of antenatal care, where there may have been a tendency to think beyond antenatal care and include care received during labour and birth. Only one woman offered a narrower definition here and she said she was only considering time spent with the midwife. Beyond this, people considered a broad and varied range of services when considering how to answer the question: for example, one woman considered a phone call to NHS Direct to have represented part of her antenatal care. This raises potential issues in terms of attributing responses to this question to particular services – additional introductory text may sometimes be needed to help focus respondents on particular services or aspects of care.

When working through the full set of questions, a number of respondents noted similarities between this and the questions on outcomes, and explicitly related their answers to the two items. Consequently we consider recommendations for these two items together, below.

## Q4. "Overall, how would you describe the outcome of...".

As in round one, the vast majority of interviewees interpreted "outcome" to refer to the clinical effectiveness of their care or treatment or, in other words, whether they had 'gotten better'. Following the first round of testing we had expressed concern that the use of the term "outcome" might be problematic in cases of continuing care; consequently we had changed the wording of the question from "rate" to "describe". Fewer such problems were noted in this round of testing, but two interviewees still did identify this as an issue.

---

- **MH** One person noted that it was difficult for them to answer the question as their treatment was ongoing: the outcome was yet to be determined.
- **GP** One interviewee told us that their last visit to the GP was simply a means of getting referred to hospital. They ticked 'good' because they got what they wanted from their visit, but did not feel that much had been done.

In the first round of testing this item had followed directly from the earlier question about experiences of care; this seemed to work well in terms of interviewees being able to note the similarities and contrasts between the questions. In the second round, though, the 'outcome' question was included second in the set of overarching items, three questions ahead of the similar question on experience. This seemed to influence people's interpretations of the item:

- **GP** One person, reviewing the questions after completion, stated that there was "no difference" between the questions on outcomes and experiences.
- **IP** Three interviewees answered the question on 'outcomes' thinking only or partly about their experiences. For example:
  - One ticked 'good', saying that although the clinical outcome of their surgery was "very good" they wanted to reflect poor aftercare in their response
  - Another answered 'good' because there had been "lots of waiting" during their time in hospital. On probing it was clear that they were in no way dissatisfied with the clinical outcome of their episode.
- **IP** Another interviewee hesitated for some time over the question. They explained that they understood the term 'outcome' to mean whether they had recovered, but assumed based on other questions that they were expected to answer based on their experiences.
- **MAT** One woman initially thought the term "outcome" included a mix of medical care, 'feelings', and 'satisfaction', and answered accordingly: however, when she later read the subsequent question asking to describe experiences she wanted to go back and reinterpret "outcome" as purely medically focussed.

Only one interviewee had serious difficulties with the question:

- **MH** One interviewee told us that they had answered the question "judging the NHS as a whole" rather than any services that they had received or the outcomes of these: however, this person made similar comments regarding other questions, suggesting a more fundamental misunderstanding.

The question appeared to be more problematic in the maternity antenatal setting ("overall, how would you describe the outcome of your antenatal care?"). Most women interpreted the questions as including the birth of a healthy baby, even though this is beyond the immediate scope of antenatal care.

## Revisions for round three

Following the second round of testing our key concern relating to this question is that the term 'outcome' may be best understood when contrasted to 'experience'. Consequently, there appears to be a good argument for aligning these items in terms of understanding – and furthermore there may be other benefits of presenting these questions together (in terms of then being able to report, contrastingly, on people's experiences and outcomes). As this affects both this and the previous item, we address changes for the two together and recommend combining them into a single two-part question for testing:

1. Thinking about <episode>, how would you describe your....

	Very poor	Poor	Average	Good	Very good
Outcome	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
Experience	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>



## Q5. “How often did your care and treatment meet/not meet your expectations?”.

This item had been heavily revised from the previous round, where it asked about best and worst care, in an attempt to make it more comprehensible and less objectionable for users. However, whilst it worked very well for around half of the interviewees, many others struggled with the new question.

The most common issue with the item was people providing an answer to only half of the question – for example ticking once to represent how often experiences met expectations but not indicating how often their experiences were not met. Only around half of the respondents placed a tick in each column as intended. Many of those answering only one column appeared visibly confused by the question, which some also expressed:

- **MH** One person said that they did not like the question as it “made you think too much” and that more intuitive answers were the preferable.
- **IP** One interviewee, who eventually answered both columns as intended, initially considered the question to be a trick – eg using scale inversion to identify inconsistent responses.
- **IP** One interviewee refused the question, asking “how can you answer two questions in one question?” – but noting that if the two questions were about specific aspects of care this would have been fine.

Conversely, many of the respondents who did answer the question successfully were very positive about it. For example, one GP patient described it as an “excellent question” that “made her think about all of her experience”. Others, however, felt that the question did not fit their experiences, which had simply met their expectation: four people stated that they would have liked to see a third column labelled “just right” or “met my expectations”.

A further issue was raised in the maternity survey, where more than one woman indicated that she did not have any expectations for her care and therefore struggled to answer the question. This seems particularly likely to be an issue for women having their first child.

### Revisions for round three

The question divided respondents, with around half not understanding it but others expressing appreciation for it. Given the high number of people answering in only one column, it seems likely that the layout may have been at fault: consequently we recommend transposing this for round three.

At this stage we do not propose to add a third row for responses labelled ‘met my expectations’. Although some interviewees requested this, we are concerned that it would make the question more rather than less complex. This will be reviewed within the round if needed.

#### 2. Thinking about all of the care and treatment you received, how often did this...

*Please tick once in each column*

	Never	Some-times	Most of the time	Always
<b>Exceed</b> your expectations	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>
<b>Fall short</b> of your expectations	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>

## Q6. “Overall...”.

As in round one, we found that the question stem of this item was unproblematic and worked well across all settings. As with the net promoter score, though, a number of interviewees objected to the use of the word ‘recommend’:

- **GP** One interviewee noted that they thought the word ‘recommend’ was inappropriate, as there was no real choice of GP practice in their area (a rural village).
- **MAT** Two women found the concept of recommending services strange. One volunteered that “you would recommend a mobile phone” but not medical care, whilst another indicated that as services were neither paid for nor ‘optional’ that it was strange to consider recommending them.

Two others were confused over the subject of the question – although both of these interviewees always experienced similar difficulties with other items.

- **MH** One interviewee understood the question to be about their recommendation of treatment, rather than the specific services they had used. They gave a positive response, explaining that they would definitely recommend that people with similar problems should seek help.
- **MH** One interviewee said that they were answering about the NHS as a whole rather than services they had used.

Encouragingly, though, other respondents had few problems with the item, and there was clear evidence that the question was being interpreted as an overarching assessment requiring people to provide a summative evaluation of the services that they had received. Sometimes this was expressed negatively, as people felt that they preferred not to assess services in this way:

- **MH** One interviewee stated that, although many aspects of their care were to their satisfaction, “not everything was good” – indicating that they considered all aspects of their experience in responding.
- **GP** One interviewee noted that it was difficult to answer the questions as there were “too many things to take into account” – they nevertheless responded with 8, which they felt gave a good reflection of a relatively mixed experience.
- **IP** One person circled 7, saying that this was a fair reflection of their hospital stay – but noting that they “didn’t like summing”.
- **GP** One interviewee circled 9, saying that they had had a very good experience overall, but noting that they disliked the scale because one “can’t quantify recommendation”.

A further issue observed during round two was reluctance from some interviewees to select the most positive option on the scale. A number of people indicated that they would not be prepared to give a rating of 10 in any circumstances:

- **MAT** One woman gave a score of 8, which she described as a good score for her as she would never give a 10.
- **MH** Two interviewees said that they would be unlikely or unwilling to give a score of 10: one said that “nothing is 100%” whilst another indicated that they “would never give anyone a 10”, because this would imply an unachievable level of perfection.
- **IP** One interviewee told us that “I never give ‘10’s [because] it’s bad luck”. Another gave a score of 9 because they said there must “always room for improvement”.
- **GP** One interviewee selected ‘9’, saying that they would not select 10 as they would not want their practice to become oversubscribed (eg because it was too emphatically endorsed and therefore became excessively popular).



**Q3.** Overall, how would you describe your experience?

4-  -Very good

2-  -Fairly good

3-  -Neither good nor poor Average

4-  -Fairly poor

5-  -Very poor

**Q4.** Overall, how would you describe rate the outcome of your <experience: eg 'visit to hospital'>?

4-  -Excellent Very good

2-  -Very good Good

3-  -Good Average

4-  -Fair Poor

5-  -Poor Very poor

**Q3 and Q4 are combined as follows:**

Q3 Thinking about the <episode>, how would you describe your....

	Very poor	Poor	Average	Good	Very good
<b>Outcome</b>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
<b>Experience</b>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>

**NB: please note that question five is considered a two part question.**

**Q5.** Thinking about all of the care and treatment you received, how often did this...

Please tick once in each column-row

	Never	Sometimes	Most of the time	Always
<b>Exceed your expectations</b>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>
<b>Fall short of your expectations</b>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>

**Q6.** Overall...

*Please circle the appropriate number*

I would not recommend <this service> I would recommend <this service>  
I had a very poor experience I had a very good experience



### 6.3. Cognitive testing interim report: round three

This report covers the third of three rounds of cognitive testing of overarching questions. This testing took place between the 28<sup>th</sup> November and 9<sup>th</sup> December 2011, and included a total of 28 interviews across four topic areas. In this document we briefly describe the range of interviews completed and questions used before looking in detail at findings around each of the six test items. Suggested revisions for round three are summarised in the enclosed appendix.

**Acknowledgment:** interviews relating to maternity services were conducted by colleagues at CQC. We are grateful for their involvement and input, and this document draws heavily on their comments when discussing findings from the maternity questionnaire.

#### 6.3.1. Interviews completed

In total, 28 interviews were conducted as part of round two.

A summary of interviews is presented in table one, below.

Topic area/ questionnaire	Number of interviews completed	Notes
GP services	8	
Acute inpatients	8	
Community mental health	10	Additional interviews were taken to correct a shortfall in the two previous rounds.
Maternity services	2	Interviews conducted by staff at CQC.
<b>All topic areas</b>	<b>28</b>	

#### 6.3.2. Questions tested

The following questions were tested in round three. Please note that the order of the questions was changed in the questionnaires in order to counterbalance for any potential order effects. For convenience, though, the questions are listed here in the same order as in the previous round one report.

**Q1.** How likely is it that you would recommend *<this service>* to a friend or colleague if they needed similar care or treatment?

*Please circle the appropriate number*

Not at all likely		Neutral		Extremely likely
0	1	2	3	4
5	6	7	8	9
10				

**Q2.** On reflection, did you get the care that mattered to you?

- 1  At all times
- 2  Most of the time
- 3  Sometimes
- 4  Rarely
- 5  Never

**NB: please note that question three is considered a two part question.**

Q3 Thinking about *<the episode>*, how would you describe your....

	Very poor	Poor	Average	Good	Very good
<b>Outcome</b>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
<b>Experience</b>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>

**NB: please note that question five is considered a two part question.**

**Q5.** Thinking about all of the care and treatment you received, how often did this...

*Please tick once in each row*

	Never	Some- times	Most of the time	Always
<b>Exceed</b> your expectations	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>
<b>Fall short</b> of your expectations	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>

**Q6.** Overall...

I had a very  
poor  
experience

I had a very  
good  
experience

0 1 2 3 4 5 6 7 8 9 10



### 6.3.3. Findings

A large number of interviews were undertaken across a range of services. Participants came from a wide range of backgrounds and had very different stories to tell in terms of their experiences. For convenience, this report summarises findings by question.

#### Q1. Net promoter score

Following round two of testing, we had adjusted the layout of this question to remove the horizontal line under the response scale. The purpose of this change was to discourage interviews from interpreting the scale as being genuinely continuous and giving non-integer responses. This change appeared to be successful as for the first time none of the interviews gave a non-integer response.

Beyond this, the item tested fairly well overall in round three, although as in previous rounds a number of interviewees still expressed their dislike of the focus on 'recommendations':

- **IP** One interviewee laughed at the question and explained that "I simply don't recommend hospitals" They nevertheless circled option 8, which seemed to reflect their stay.
- **IP** One interviewee seemed confused by the question, asking "why would I recommend bladder cancer?". They nevertheless circled option 8, which seemed to reflect their stay.
- **MAT** A respondent asked 'whether anyone "recommends" the NHS' as it is 'just there for everyone'. This person went on to interpret the question as being about the choice between NHS and private care, rather than her particular provider.

Several interviewees interpreted the question as being recommending seeking treatment rather than using a specific service:

- **IP** One person noted that the question was odd because "if someone was ill you'd say 'go to hospital'".
- **MH** One respondent ticked 9, explaining that it was "better to do something than nothing". They further indicated that they would have ticked a 10 but that "some people may be reluctant" to seek help for a mental health condition. The very positive response circled was not consistent with their other answers; answering another question this person described their treatment as "slightly above average".
- **MH** One interviewee ticked 10, stating that as their treatment was free "it would be silly not to recommend it". Their own experience was not so positive, though, and the response of 10 certainly did not seem to reflect their view of their care.
- **GP** One interviewee ticked 10 for this question, despite giving far less positive responses elsewhere and noting serious concerns about their treatment by a practice nurse.
- **MAT** Another interviewee was unsure what the question was asking about, and noted that it was "weird". She considered that the question was asking if she would recommend a water birth, as she had had, rather than the hospital where this took place.

Finally, some linguistic issues were noted:

- **GP** An interviewee whose first language was not English had difficulty with the question, particularly with interpreting the word "neutral", which they repeatedly read aloud when considering the question.

## Q2. “On reflection, did you get the care that mattered to you?”

As in the first two rounds of testing, interviewees gave very mixed reactions to this item. Many were able to respond quickly and gave clear, reasoned explanations of how they had arrived at their answer. Some answered with regard to numerous aspects of care, whilst others picked the one factor that mattered most to them. In most cases, people described relational aspects of care, but a number of interviewees answered purely in terms of clinical outcomes: for example, one hospital inpatient who had been treated for cancer said that what mattered to them was “being in remission”.

In this round of testing a number of people expressed a dislike of the question:

- **GP** One interviewee found the question inappropriate and said that “the care that ‘mattered’ is irrelevant [because] sometimes you need care that you don’t want”.
- **MH** One interviewee remarked that the question was “too subjective” and that they didn’t understand what it meant.
- **MH** Another interviewee said that the question “seemed pointless”.

It is notable that the question seemed to raise far more objections in this round, when placed last in the set of overarching questions, compared to in the previous round where it was situated at the front of the set. As in previous rounds, interviewees expressed a concern that the question may be particularly susceptible to order effects.

## Q3. “Overall, how would you describe your experience” and

## Q4. “Overall, how would you describe the outcome of...”.

In previous rounds we found that the ‘overall, how would you describe your experience’ item worked well, albeit with some concerns about the explicit use of the word ‘experience’ in the question stem. Likewise the comparable question on outcomes worked well and respondents often explicitly related their answers to these two items. Consequently for the third round we tested these two items as a combined, two part question.

In this round of testing, we found the main issue associated with these questions to be the layout. Five respondents (three in the GP setting and two in the inpatient setting) ticked only in the top row before moving on without answering the second part of the question. This problem seemed particularly pronounced amongst people for whom English was not their first language, and those with lower literacy.

By contrast, the layout worked well for many other respondents. The majority of people interviewed gave different answers regarding their experiences and outcomes or, if ticking the same option for both, gave clear and distinct explanations of their reasoning for each case. The vast majority of interviewees understand the terms ‘outcome’ and ‘experience’ clearly, and positioning these items together seemed to encourage respondents to interpret them in distinct ways. For example, one interviewee said that the difference between ‘experience’ and ‘outcome’ was that they reflected the short term versus the longer term: that is, what the *experience* of the appointment and consultation was like versus the eventual *outcome* in terms of an improvement in their health. Only one person, an interviewee in the maternity setting, told us that ‘outcome’ and ‘experience’ “meant the same thing”. There were, however, still issues with these terms in a minority of cases:

- **MH** One person struggled to answer about their outcome. They understood the terminology well, but said that although their mental health had improved this was not because of the care they had received. They eventually ticked ‘average’ for this row.

- **MH** Another interviewee told us that they didn't understand the term "outcome", which they described as "jargon". When prompted, they were able to accurately guess what the term meant.
- **MAT** One woman struggled with the term 'outcome' (in relation to antenatal care) as she was unsure what it was referring to. She asked if it meant whether she was able to give birth, but skipped the question.

This latter issue reflected similar problems in previous rounds, where people interpreted the outcome of all maternity care – including antenatal services – as being the birth of their child.

## Q5. "How often did your care and treatment meet/not meet your expectations?"

This item was revised for round three, with the response options transposed into rows instead of columns. Unfortunately, this did not successfully address the problems found in the previous round.

Once again, the most common issue with the item was people providing an answer to only half of the question – for example ticking once to represent how often experiences met expectations but not indicating how often their experiences were not met. Only around half of the respondents placed a tick in each row as intended, and several appeared confused:

- **MH** One person noted that the question felt confusing because the order of the scales seemed to swap: a tick in the leftmost box was positive on one row and negative on the other. This person did, however, understand the question well and answer without difficulty.
- **IP** One interviewee struggled to answer the question properly, ticking 'never' in both rows to try to indicate that they had had a very bad experience.
- **GP** One interviewee ticked every box on the first row but none on the second.

Some interviewees also objected to the use of the term 'expectations':

- **IP** One interviewee disliked the word 'expectations', saying that when going to hospital "the aim is just to get better". Oddly this person did not see 'getting better' as an 'expectation' *per se* – perhaps more as an objective.
- **GP** One interviewee answered the question as intended but said that it was "odd" and that they "hated it" because "expectations don't matter".
- **MAT** One interviewee wondered whether the question would be appropriate for people who might have unrealistic expectations: eg if someone had written a birth plan for themselves that was not medically appropriate.

Overall, there seemed to be numerous remaining issues with this question, and it was arguably the least successful of the items tested in this round.

## Q6. "Overall..."

Following round two of testing, the response labels for this question were changed to focus on 'experiences' rather than 'recommendations'. This appeared to work well in that interviewees tended not to express concerns about the nature of the question, with only one exception:

- **MH** One interviewee stated that having a mental health condition was clearly not a "good experience" in and of itself. Nonetheless, they understood that the question was asking about the services they received and ticked 10, which was consistent with their responses elsewhere and what they described as a very positive experience of care.

In general, this question worked very well in this round of testing and raised few concerns. Only one interviewee was confused about the question: a respondent to the GP survey read the response scale and asked “experience of what?”. This person was, however, still able to answer the question appropriately, and it should be noted that they had difficulties with written English and struggled with many items in the questionnaire.

The only other issue noted in this round regarded the use of the response scale. As in previous rounds, a minority of interviewees gave non-integer responses (an inpatient selecting 8.5 and a GP patient selecting 9.5). As both of these interviewees went on to give integer responses for the subsequent net promoter score question, where the line beneath the response options had been removed, it is possible that removing this line from the “Overall...” question would be beneficial.

## 7 References

---

- i Tourangeau, R., Cognitive sciences and survey methods. In T. Jabine, M. Straf, J. Tanur, & R. Tourangeau (Eds.), *Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines*, pp. 73-100. Washington, DC: National Academy Press
- ii Reichheld, F.F. (2003). The One Number You Need to Grow. *Harvard Business Review*, 81, 46-54.
- iii eg Reichfeld, F.F., Markey, R. (2011) *The Ultimate Question 2.0: How Net Promoter Companies Thrive in a Customer-Driven World*. Harvard Business Review Press: Boston, MA.
- iv Keiningham, T.L., Cooil, B., Andreassen, T.W., & Aksoy, L. (2007). A Longitudinal Examination of Net Promoter and Firm Revenue Growth. *Journal of Marketing*, 71(3), 39-51.
- v Kristensen, K., & Eskildsen, J. (2011). Is the Net Promoter Score a reliable performance measure? *2011 IEEE International Conference on Quality and Reliability*, pp249-253.
- vi Schneider, D., Berent, M., Thomas, R., & Krosnick, J. (2008). Measuring customer satisfaction and loyalty: Improving the 'Net-Promoter' score. *Poster presented at the Annual Meeting of the American Association for Public Opinion Research*, New Orleans, Louisiana.
- vii Nuyens, R. (2011) *The Effects of Negative Celebrity Endorser Publicity on Consumer Attitudes and Behavioural Intentions*. Maastricht University of Business and Economics.
- viii van der Schee, E., de Jong, J. D., & Groenewegen, P. P. (2011). The influence of a local, media covered hospital incident on public trust in health care. *The European Journal of Public Health*. doi:10.1093/eurpub/ckr033
- ix *op cit*
- x eg [http://www.northwest.nhs.uk/document\\_uploads/Board%20papers%20September%202010/15.%20INSPIRE%20update.pdf](http://www.northwest.nhs.uk/document_uploads/Board%20papers%20September%202010/15.%20INSPIRE%20update.pdf)
- xi Wrightington, Wigan, and Leigh NHS Foundation Trust. (2011) *Advancing Quality Progress Report*. Draft document published at [http://www.wvl.nhs.uk/Library/Trust\\_Board/FT\\_Mins/2011/November/Advancing\\_Quality\\_Progress\\_Report.pdf](http://www.wvl.nhs.uk/Library/Trust_Board/FT_Mins/2011/November/Advancing_Quality_Progress_Report.pdf), last accessed 13/01/2012.
- xii Tourangeau, R., Rips, L.J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge University Press: Cambridge, England. p45
- xiii eg Knauper, B., Belli, R.F., Hill, D.H., & Herzog, A.R. (1997). Question difficulty and respondents' cognitive ability: the effect on data quality. *Journal of Official Statistics*, 13(2), 181-199.
- xiv Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied cognitive psychology*, 5(3), 213–236.
- xv eg Dillman, D. A., Smyth, J. D., & Christian, L. M. (2009). *Internet, mail, and mixed-mode surveys: The tailored design method*. Wiley: NJ, US.