NHS England

Review of the Friends and Family Test

NHS England INFORMATION READER BOX		
Directorate		
Medical	Operations	**Patients and Information**
Nursing	Policy	Commissioning Development
Finance	Human Resources	

Publications Gateway Reference:	**01832**
Document Purpose	Guidance
Document Name	NHS England Review of the Friends and Family Test
Author	NHS England
Publication Date	21 July 2014
Target Audience	Directors of Nursing, NHS England Regional Directors, NHS England Area Directors, NHS Trust Board Chairs, Directors of Finance, NHS Trust CEs
Additional Circulation List	
Description	The Review was designed to assess the implementation, publication, use and reception of FFT during the first six months of national data collection, focusing on what has worked well and what might be improved. The evidence gathering process was begun in November 2013 and was completed in March 2014.
Cross Reference	N/A
Superseded Docs (if applicable)	N/A
Action Required	For information
Timing / Deadlines (if applicable)	
Contact Details for further information	Dan Wellings Head of Insight & Feedback NHS England 6th Floor, Skipton House 07803 085617

Document Status

This is a controlled document. Whilst this document may be printed, the electronic version posted on the intranet is the controlled copy. Any printed copies of this document are not controlled. As a controlled document, this document should not be saved onto local or network drives but should always be accessed from the intranet

Review of the Friends and Family Test

Inpatient and Accident & Emergency settings

Version number: 1

First published: 21st July 2014

Prepared by: Insight Team, NHS England

Contents

1 Overview

The Friends and Family Test (FFT) was introduced in April 2013 so that patients visiting A&E or staying in hospitals as an inpatient could rate and comment on their experience. The prime purpose of FFT was to make this feedback publicly available so that other patients and the public at large could use that information in making their own decisions and choices about hospital care. This review, based on thorough research and analysis, has looked at the impact of FFT in its first year of operation. It has found that, in addition to its ambition to provide data to inform the public, FFT has proved to be a powerful tool for service improvement by promoting a culture of increased responsiveness to patient feedback in the NHS. Through FFT, NHS staff are receiving regular, near real time feedback about the job they are doing. Often this feedback confirms what a great job they are doing, but where it is less positive it encourages staff to make changes in order to improve the quality of care experience.

This review presents evidence concerning the first six months of FFT in the A&E and inpatients settings. It focuses on the implementation, publication, use and reception of FFT, as well as assessing its methodology and the quality of its data. It is informed by large-scale independent qualitative research with NHS staff, stakeholders, patients and the public, as well as quantitative analysis of existing FFT data. In addition to describing FFT's performance thus far, it highlights examples of best practice and makes recommendations designed to promote the effectiveness and value of FFT for the NHS and the public.

The subsequent sections of the review cover the following:

- **Introduction:** Explanation of the context of FFT and of the review methodology
- **Ambitions of FFT:** Description of how FFT is performing against its original ambitions
- **FFT data:** Analysis of the quality and uses of the data produced by FFT
- **The FFT Score:** Assessment of the effectiveness of the current FFT score and a consideration of alternatives
- **FFT as a local feedback mechanism:** A focus on FFT's potential for effecting service improvement and cultural change in the NHS
- **Publication of FFT data:** Consideration of challenges and opportunities for the national and local publication of FFT data
- **Conclusions and Recommendations:** Our suggestions of ways of modifying FFT to make it as effective as possible going forward.
- **Appendices:** The evidence that this overall report is based on: independent qualitative research, a review of FFT data quality and a paper on the suppression of data.

Equality and diversity are at the heart of NHS England's values. Throughout the development of the policies and processes cited in this document, we have given due regard to the need to eliminate discrimination, harassment and victimisation, to advance equality of opportunity, and to foster good relations between people who share a relevant protected characteristic (as cited in under the Equality Act 2010) and those who do not share it.

2 Introduction

The introduction of the Friends and Family Test (FFT) was announced by the Prime Minister in a speech in May 2012.

> *"In every hospital, patients are going to be able to answer a simple question: whether they'd want a friend or relative to be treated there in their hour of need. By making those answers public we're going to give everyone a really clear idea of where to get the best care - and drive other hospitals to raise their game."*[1]

Since April 2013, the FFT has been implemented by all providers of NHS funded acute services for inpatients[2] and in Accident and Emergency (A&E) departments.[3] The guidance for implementation of FFT was published by the Department of Health on 4th October 2012. As of 1st October 2013 FFT was extended to include all women of any age who use NHS funded maternity services, and on 1st April 2014 FFT was launched with NHS staff. It will be rolled out to GP, community and mental health services by the end of December 2014, with outpatient and day cases, ambulance services and dentistry to follow by the end of March 2015.

The test asks a single question: 'How likely are you to recommend our <ward / A&E department> to friends and family if they needed similar care or treatment?' Patients choose a response on a five point scale from 'Extremely unlikely' to 'Extremely likely' (or can select 'Don't Know'). The responses are then combined to give an overall FFT score, by subtracting the proportion who are 'Extremely unlikely,' 'Unlikely' or 'Neither likely nor unlikely' to recommend from those who are 'Extremely likely' to recommend. The question and score calculation are modifications of the 'Net Promoter Score' (NPS) methodology, which is widely used in the private sector.

The NPS was developed as a measure of brand loyalty by Fred Reichheld, Bain & Company, and Satmetrix. It asks customers whether they would recommend a product or service to their friends and family.[4] Like FFT, it classifies responses to its question as 'Detractors,' 'Neutral' and 'Promoters,' and presents a net score of promoters minus detractors. Unlike FFT, however, the classic NPS asks customers to respond on a scale of 0-10, is often asked at several stages of a customer journey rather than after a single event, and the score is used almost exclusively for internal improvement and incentivisation rather than for external communication.

---

[1] https://www.gov.uk/government/news/friends-and-family-test-aims-to-improve-patient-care-and-identify-best-performing-hospitals.

[2] Including independent sector organisations that provide acute NHS services.

[3] More specifically, it is asked of patients discharged from Type 1 and Type 2 A&E departments:
Type 1 A&E department (Major A&E): A consultant led 24 hour service with full resuscitation facilities and designated accommodation for the reception of accident and emergency patients;
Type 2 A&E department (Single Specialty): A consultant led single specialty accident and emergency service (e.g. ophthalmology, dental) with designated accommodation for the reception of patients.
FFT has not been implemented in other types of A&E department, minor injury units (MIUs) or Walk-in Centres (WiCs) (Type 3).

[4] A registered trademark of Fred Reichheld, Bain & Company and Satmetrix, NPS was introduced by Fred Reichheld in a 2003 journal article: Reichheld, Frederick F. (December 2003). 'One Number You Need to Grow".' Harvard Business Review.

The FFT joins existing patient experience measures in NHS provider settings. The CQC runs an extensive survey programme, for example since 2002 the annual NHS inpatient survey has provided a detailed and representative measurement of the experience of inpatients at each hospital trust. In addition, many trusts carry out their own work, both qualitative and quantitative, into patient experience.

2.1 Purpose of this review

NHS England made a public commitment to review the methodology and implementation of FFT after the first 6 months of data collection in the inpatient and A&E settings in order to identify possible improvements. The evidence gathering began in November 2013 and was completed in March 2014. This report presents the findings of this work.

NHS England's review of FFT had the following specific purposes:

- To assess the implementation, publication, use and reception of FFT during the first six months of national data collection, focusing on what has worked well and what might be improved;
- To collect and present evidence of best practice across the settings;
- To judge whether the test in its present incarnation matches its original ambitions, and, if necessary, to highlight any changes that would be required to fulfil all of these functions, specifying any associated risks or costs; and
- To ensure that future NHS England guidance on FFT is evidence-based and reflects best practice.

2.2 Methodology

The evidence for the review was gathered via two main strands of research, summarised below: (a) qualitative evidence gathering (with a quantitative element), and (b) quantitative analysis of FFT data:

2.2.1 Qualitative evidence-gathering with trusts and stakeholders

The qualitative strand of the review ('the qualitative review') was conducted by Ipsos MORI, the independent social research agency, following a competitive tender process. Their research focused on the implementation, reception and use of FFT in clinical settings (using structured feedback from 95 of 156 NHS trusts – a response rate of 61% - and in-depth case studies of nine trusts), the reception of FFT by patients and the public (through ten focus groups), and the views of experts and stakeholders about the first six months of FFT (through twenty in-depth interviews).

2.2.2 Quantitative data analysis

Quantitative analysis of FFT data ('the quantitative review') was conducted by NHS England analysts, working in collaboration with the Office for National Statistics (ONS), who independently peer-reviewed the analysis and provided quality assurance. The analysis investigated the inter-trust and inter-ward comparability of FFT data, effects due to differences in data collection mode, and tested the current FFT scoring metric against possible alternatives.

The independent Ipsos MORI qualitative research report ('the qualitative review'), and the NHS England quantitative analysis peer-reviewed by ONS ('the quantitative review'), are attached as appendices to this current report. In the writing of this overall summary report, additional literature relevant to FFT was consulted, including research conducted during the FFT design and implementation process prior to April 2013.

3 Ambitions of FFT

This section describes the four original ambitions for FFT, before reviewing how well it is currently performing in relation to each of these different functions. This evidence has three sources: an online feedback form completed by 95 of the 156 trusts that it was sent out to (a response rate of 61%); case study visits to 9 trusts, where the use of FFT was observed and interviews and focus groups were conducted with staff, local patients and members of the public, and; 20 in-depth interviews with stakeholders.[5]

3.1 The ambitions of FFT

The original ambitions for FFT were as follows:

- A tool for local service improvement
  - Using patient feedback to inform the way services are delivered by understanding where and how improvements can be made

- Providing information to empower patient choice
  - Enabling the public to use FFT feedback to inform their decision about from which hospital or trust they wish to receive their treatment or care

- Intra-trust performance monitoring
  - To provide patient experience information at ward level capable of differentiating performance within a hospital, highlighting areas of best practice and areas of concern

- Providing comparable data to hold trusts to account
  - Using FFT scores as a way of identifying and rewarding better performing trusts

3.2 Performance of FFT against its ambitions

3.2.1 A tool for local service improvement

3.2.1.1 Overall findings

The research found that FFT is performing well as a service improvement tool, with 85% of trusts reporting that it is being used to improve patient experience, and 78% saying that FFT has increased the emphasis placed on patient experience in their trusts.[6] In particular, the open-ended follow-up question is proving a rich source of patient views at ward and A&E departmental level that can be used, where implemented, to highlight best practice and identify and address concerns.

[5] This evidence is presented in greater detail in Ipsos MORI's research report, which is appended to this document.
[6] Ipsos MORI (2014) *The Friends and Family Test: Qualitative research*: pp.1-2. Research conducted on behalf of NHS England.

While this is a positive finding, highlighting the strong potential of FFT going forward, not all trusts are currently deriving full value from the FFT. In particular, some trusts do not currently ask the follow-up question about why a patient gave a particular response, which necessarily limits the extent to which staff will understand what they are doing well and what needs to be addressed and improved. Also, the research found that not all staff are equally engaged in the FFT data, with clinicians tending to be much less aware of FFT, particularly in the inpatient setting.

There are varying levels of staff exposure to FFT results. In many trusts the FFT scores and comments are displayed on ward noticeboards, where frontline staff, along with patients, their friends and family, can view monthly results. Often these results are displayed alongside other patient experience and quality measures, such as local survey results or metrics on pressure ulcers and falls, and in some cases a sample of qualitative FFT comments are presented alongside the quantitative FFT data. In some trusts, however, staff had limited or infrequent access to FFT scores and comments, which has a clear impact on their ability to respond to the feedback.

FFT's capability for delivering real-time feedback was found to be a particular strength for its use in local service improvement, as opposed to more rigorous and representative methods of measuring patient experience. The National Inpatient Survey,[7] for example, takes approximately six months to return its data, and does not report at ward level. As such, the overall findings of the data are not especially salient for frontline staff, since they cannot be sure that they reflect the care they provided, and they may find it difficult to relate findings to events from several months earlier. FFT, by contrast, provides ward specific data, and does so soon after the event. Staff are thus able to associate the feedback with what was happening on the ward at the time: both of these aspects were found to increase the relevance of the information and hence the motivation for staff to digest and respond to it.

In the qualitative research with stakeholders, there was general agreement amongst representatives of organisations such as the CQC, Monitor, the Royal College of Nursing, the Foundation Trust Network and the Department of Health that FFT is a useful new tool in hospitals. They saw FFT as being instrumental in introducing a new culture: one in which patient feedback is more swiftly listened to and acted upon. Some went so far as to suggest that FFT put 'the patient at the heart of everything.' The evidence from this review suggests that this cultural change has begun, but that there is some way to go before it is happening uniformly across the NHS. It is clear the FFT is performing well as a tool for service improvement and that this promise should continue to be built upon. How FFT can best be utilised for service improvement is discussed in section 5.

3.2.1.2 Use of real-time feedback

Although FFT was created with a number of ambitions in mind, frontline staff made clear in discussions that they primarily understood FFT as a local service improvement tool. It proved most effective when clear and actionable feedback was

[7] The National Inpatient Survey, conducted by the Picker Institute Europe on behalf of the Care Quality Commission. http://www.cqc.org.uk/public/reports-surveys-and-reviews/surveys/inpatient-survey-2013.

communicated to frontline staff in a timely manner so that they could respond quickly to perceived issues or complaints. Additional benefits included the use of positive feedback to motivate and encourage staff.

Staff felt that, considered on its own, the FFT score was of limited value. As is discussed in the section 4, below, the overall score at ward level proved somewhat difficult to interpret as a measure of patient experience. Likewise, individual responses to the main FFT question, considered without explanatory feedback from the follow-up question, sometimes appeared abstract to staff, as it was not clear which aspects of the patient's experience informed the rating they gave.

Frontline staff were most enthusiastic about FFT where a follow-up question was being used. The open-ended nature of the question allows patients to provide additional feedback in their own words, and in any way they wish, contextualising the FFT data and delivering insight. This produces a variety of comment types, from describing general sentiment to mentioning specifics about the experience.
Much of the follow-up feedback to FFT is positive, often praising staff or highlighting positive aspects of the patient's stay. The interviews with Patient Experience leads and nursing staff highlighted the degree to which such positive feedback can be used to encourage staff and to reinforce good practice. In particular, comments that acknowledged good care or which praised the dedication of the nursing staff were found to be highly motivating. Staff might not otherwise receive direct praise or encouragement from the patients in their care, and some commented that it provided a very useful antidote to some of the more negative stories in the media about NHS workers, ultimately making them feel more valued. One nurse described the effect this has had:

> *"In the press, the perception is that nurses aren't caring anymore, but I would say the FFT has reassured me that patients do think we're kind and caring."*[8]

Irrespective of the score given, arguably the greatest practical benefit delivered by FFT came when patients went on to identify improvements that could be made. This was especially the case when patients explained the reasons for their negative or neutral FFT responses, helping staff to identify issues and resolve them, or at least improve the situation.

In many cases, a patient who answered the question with a positive response overall still chose to comment on something that could be improved and this aspect of FFT was arguably found to be providing the most practical benefit to the service.

While some FFT comments simply helped staff understand more about the patient experience of their care, causing them to reflect on their approach, other comments highlighted specific, discrete issues that could be definitively addressed. In some cases these were problems that staff had been unaware of, while in other cases the FFT feedback corroborated previous suggestions for improvement that had not been approved and staff were able to present the feedback to their management as evidence to convince them of the benefit of acting. Many of the anecdotal

[8] Ipsos MORI (2014) *The Friends and Family Test: Qualitative research*: p.94. Research conducted on behalf of NHS England.

improvements prompted via FFT were found to be environmental, such as improved cleanliness, warmer water in showers, or noise reduction through the purchase of silently-closing bins. Other improvements were more service orientated, such as the provision of better choice or quality of meals. Some feedback prompted actions to change the behaviour of staff, such as the reduction of noise at night. In other cases, however, FFT had prompted creative solutions to more complicated issues, such as the creation of a new 'Discharge Coordinator' role in order to address the problem of delays to patient discharge, or the introduction of "You missed us" cards to notify patients who were not present in their room to receive their pain relief.

3.2.1.3 Variation in use of FFT among trusts

Research with trusts found significant variation in the way FFT feedback was disseminated and responded to by staff. Some ward managers were able to view the feedback several times a week and communicate the findings to staff, whereas others only received a monthly summary of the data. The quality of the FFT data made available to staff differed greatly in quality: while some reports analysed the collated data in various ways and included all qualitative feedback, other trusts produced nothing more than the nationally-mandated submission. Likewise, the way in which FFT feedback was tailored for staff varied a great deal, with some wards able to view their own data in detail and in comparison to other wards in their hospital, while in other cases ward staff were unaware that ward-level data was available. Also, the use of FFT data on the front line varied greatly, with some wards discussing the feedback at weekly meetings and planning responses as a result, while in others FFT was not routinely discussed amongst staff or communicated on ward noticeboards.

Three factors appeared most influential for how FFT was working as a local feedback mechanism:

- the culture and infrastructure of patient experience data that existed prior to FFT;
- the leadership shown in the trust on implementing FFT, and on learning from FFT and other patient experience data;
- the solution to data collection and processing that the trust or department had implemented, particularly the mode of data collection used, and whether or not a supplier had been engaged.

A trust's capability and resource to consistently and effectively act on feedback was found to be an important factor in determining the local success of FFT. Where patient experience data was already routinely gathered and viewed by staff, such that there was an existing culture of continual improvement, FFT tended to be more integrated in the working routine of the ward, with staff more likely to value patient experience data and to have viewed the latest FFT feedback. This often occurred where a specialist role of Patient Experience Lead already existed within the trust. This suggests that, in wards and departments that previously lacked a culture of collecting and responding to patient experience data, FFT may take longer to become embedded and used as effectively as possible. Trusts that have until now

placed low emphasis on patient feedback may in particular benefit from considering best practice models of data usage.

The overall leadership on FFT and patient experience data more generally was found to influence the extent to which FFT was being used to drive service improvement. Achieving the CQUIN on response rates was found to have been the main area of focus for many trusts in the first 6 months of FFT operation.[9] In some cases, this meant that greater emphasis had been placed on response rates than on using the data. Discussions with frontline staff in the case study trusts showed there were differing levels of engagement with the data (both the score and the qualitative feedback). Some hospitals were being encouraged to use the comments to form action plans for improving patient experience, whereas others were predominately focused on achieving the 15% response rate.

The solution to FFT data collection and processing employed by a trust determined many of the options for how the feedback could be used by staff. Some solutions allowed for an immediate view of the data, whereas others produced only a monthly collated view. For example, where data was being collected by staff on postcards, ward managers often had the opportunity to review the cards at the end of a shift in order to get an overview of the feedback, albeit in a piecemeal fashion, and to identify any immediate issues of concern. In some trusts where data was being collected electronically, FFT feedback was immediately uploaded to the staff intranet. The most sophisticated systems collated the data in a variety of formats, such as over time or by ward, and also allowed for the feedback to be reviewed at response level, including the free text comments. In other cases, such as where a supplier was collecting the data from patients after discharge (e.g. by SMS, telephone call or internet survey), staff were not be able to view the feedback until the monthly collation had taken place.

In some cases, electronic forms of data collection allowed free text comments made by patients to be reviewed and more easily analysed. Some trusts were beginning to employ approaches to analysing the qualitative feedback received in order to detect trends in the data, looking at common themes or using sentiment analysis. However, where paper-based data collection methods were being used, this type of computational analysis first requires the free text comments to be electronically scanned or transcribed. Not all paper-based trusts were found to be currently doing this, meaning that their ability to analyse the patient comments was restricted. One advantage of paper-based methods was that at the end of each shift staff could read feedback, which arguably makes it more impactful, but a corresponding disadvantage is that patterns in the comments over time are more difficult to discern. Also, not having the patient comments available on the staff intranet meant that fewer frontline staff had the opportunity to read the comments, since the physical postcards or questionnaires would not be widely shared.

[9] A small percentage of hospital providers' payments is determined by their performance, as specified by the Commissioning for Quality and Innovation scheme (CQUIN). In the financial year 2013-14, national CQUINs for FFT included payments for (a) achieving a 15% response rate for FFT in Q1 in inpatients and A&E, and (b) for achieving an improved response rate of at least 20% in Q4. See: http://www.england.nhs.uk/wp-content/uploads/2013/02/cquin-guidance.pdf. For further clarification see also: http://www.england.nhs.uk/wp-content/uploads/2013/07/cquin-faq.pdf.

The communication of FFT scores to staff was inconsistent across the trusts included in the case study element of the review. In some cases scores were displayed widely around the hospital, in public areas and on ward noticeboards. All-staff emails and the intranet provided other ways to disseminate FFT data. In other trusts FFT was far less visible, and this had a noticeable impact on staff knowledge and comprehension, and arguably reduced its impact as a patient experience measure. Amongst trust staff, doctors tended to be the least aware of FFT data, as they were less exposed to the collection and its subsequent use on wards.

A further difference in approach concerned the nature of the follow-up question. As has been discussed, in some cases follow-up feedback was not always collected, especially where a token system was in use. Where it was collected, the review encountered significant variation in the wording of the follow-up question. Different variants asked for an explanation of the patient's response to the FFT question, for what was best/worst about their stay or care, or for what the ward could improve on. We would expect slightly different emphasis from the feedback depending on what is asked.

3.2.2 Information to support patient choice

The qualitative review found low awareness of FFT among the public, such that few would think to use it as part of a decision about which services to use. The phrase, 'Friends and Family Test' was not spontaneously recognised in the discussion groups, although when presented with the question wording some recent patients thought they had answered the question during a recent stay in hospital or A&E.

When they were formally introduced to the concept of FFT, patients and the public tended to initially think of it as a measure that governing bodies would use to hold hospitals to account, or as something each trust would use in order to monitor its own performance. On reflection, after being told that FFT results are publicly available, participants understood how FFT could be used to compare different hospitals, and some were positive about its ability to support choice. However, choice was felt to be more relevant in some clinical settings than others. In the case of an emergency, for example, many felt that choice would be irrelevant as they would most likely simply go to their nearest A&E. Some also pointed out that choice was not always offered to patients. It was also felt that publicising ward-level FFT data was not helpful, as patients would be highly unlikely to be able to choose the ward they were going to be treated on. Nevertheless, public and patients were positive about the collection of patient experience feedback and expressed an interest in having access to FFT data in order to support their decision-making.

Although the possibilities of informing choice were viewed favourably by patients and the public, they felt that the main way of presenting the FFT data was not a useful summary statistic or measure. In its current form, patients and the public found the FFT 'net promoter' score (of 'promoters' minus 'detractors') to be difficult to understand and interpret. The research found a strong appetite for an FFT score that is simpler and easier to understand, and many participants expressed a wish to be able to explore the full set of FFT responses for a hospital, and to have access to patient comments.

NHS staff expressed reservations over the ability of FFT to inform patient choice, citing data comparability issues as well as difficulties in interpreting and understanding the score. Stakeholders, particularly those with a survey methodological background, were of a similar view. When trust staff were asked about the patient and public knowledge of FFT there was a widely held perception that relatively few patients were currently using the data.

Commissioners also suggested that they do not think FFT is being used to inform patient choice, although this is as much based on the fact that they do not see choice being offered, or exercised, rather than any comment on the ability of FFT to perform this role.

3.2.3 Intra-trust performance monitoring

Within trusts, FFT was being used by senior management to monitor ward performance, although many Patient Experience leads and frontline staff recognised that the variation in ward specialism and size can make comparisons difficult. There was also an acknowledgement that some wards, such as those specialising in care of the elderly, represented a particular challenge in terms of collecting the data, resulting in lower response rates and a corresponding lack of representativeness.

Board level staff explained that they would occasionally visit wards, where they would be able to view FFT results. In the main, though, their primary interaction with FFT was in monthly or quarterly board reports. In most cases a range of patient experience measures, including FFT data, were reported to the Board on a monthly basis. Here, scores at ward and trust levels were presented alongside other patient experience data in summary dashboards. In some cases the addition of verbatim feedback from the follow-up question provided supplementary information alongside which the score could be viewed.

Intra-trust comparability of FFT scores was generally regarded as a valid exercise, with management using ward-level data to identify poorly performing wards or sites. Given consistency in approach to FFT within trusts, the monitoring of ward and departmental performance over time also appears a potentially valuable use of FFT, particularly as an early-warning system about areas that are showing poor performance. Further consideration needs to be given to this and how using the data this way can work for national and local bodies.

3.2.4 As a transparent headline metric of performance which can be used to hold trusts to account

The role of FFT in the performance management of trusts is currently relatively limited. Thus far the CQUIN payments attached to the monthly data collection have focused on response rates. A reasonably high response rate is viewed as one way of ensuring that aggregated FFT results possess validity, providing a reasonable representation of patients' views. While CCGs may choose to employ FFT scores within local CQUINs, we understand that few, if any, are currently doing so. Future decisions about the suitability of FFT for discriminating between the performance of trusts, and of rewarding trusts accordingly, may be based in part on the analyses presented in this review.

In the small number of interviews conducted with commissioners, the qualitative review found that, as a result of the CQUIN, their main interest in FFT was the response rate for each trust, and whether the payment threshold was being met. The six commissioners interviewed were uncertain about the value of the FFT score as a standalone measure, and suggested that FFT data would need to be triangulated with other patient experience metrics, such as complaint volumes, in order to contribute to a reliable measure of trust performance. In some trusts the FFT is part of the regular contract monitoring conducted by CCGs. Some question the validity of FFT to add value to these discussions.

More generally, however, the qualitative review encountered a positive response to the introduction of FFT from commissioners. FFT is seen as a way of further embedding patient experience measures in the provider landscape, and patient experience measures form part of the regular conversations between CCGs and providers. Most of the six CCGs who took part in the qualitative review received quarterly reports from providers, presenting the headline FFT scores and the response rates over each period. There was also a desire to view the open-ended comments collected via the follow-up question as well, but this is not routinely happening.

In the online survey that was completed by trusts, around three quarters said they had used the FFT to compare the performance of their trust with others.[10] The qualitative review found that senior staff in particular would sometimes use FFT scores to compare performance against other nearby or similar trusts, though most expressed scepticism about the validity of such comparisons, given the differences in approaches to data collection, and variations in response rates.

Other organisations are including FFT as part of their health service monitoring process, including the Care Quality Commission (CQC). As part of its safeguarding responsibilities the CQC maintains an 'intelligent monitoring system'. This looks at a range of published measures to identify and target investigations of healthcare providers on the basis of risk. At present CQC includes the inpatient FFT scores in this system, but decided to remove the A&E FFT from its analysis, due to question marks over reliability of the data in this setting.

[10] Ipsos MORI (2014) *The Friends and Family Test: Qualitative research*: p.87. Research conducted on behalf of NHS England.

4 FFT data

This section presents evidence about the characteristics of FFT data, and what this means for how FFT data is used. In particular, it discusses whether FFT data should be used to compare the performance of different trusts, and the advantages and disadvantages of standardising the way FFT data is collected.

The Department of Health implementation guidance for FFT presented a permissive methodology with regard to data collection, which allowed trusts to gather data by any appropriate means. All eligible patients should be offered the test, and the test should be administered at the point of discharge or within 48 hours of discharge, but the mode of data collection is not specified. As such, trusts – and in some cases different wards within trusts – have developed a range of solutions to collecting FFT data. In many instances a range of techniques are employed within the same site, and data from one site is often a mixture of responses given at the point of discharge and responses provided up to 48 hours after discharge.

4.1 Response rates

The response rate of a survey refers to the proportion of eligible participants that complete the survey. The most common reasons for non-response are failure to receive the survey ("non-contact"), refusal to complete the survey, and incapacity to complete the survey. In this section we discuss the effect the response rates for FFT have on how we can use the FFT data.

4.1.1 Overview

The mean response rates for the FFT are generally lower than for national health surveys in similar settings, although much variation is observed between different sites. In February 2014, the inpatient response rate was 34.2% (ranging from 100% to 0% at site level) and the response rate in the A&E setting was 18.6% (ranging from 89.2%[11] to 0% at site level).[12] By comparison, the 2013 National Inpatient Survey achieved a 49% response rate,[13] and the 2012 National Accident and emergency department survey achieved a response rate of 38%.[14] Many factors may contribute to the lower response rates of FFT, such as the fact that FFT is usually only offered once to individuals, whereas reminder letters or additional questionnaires are commonly distributed to survey non-respondents. The point of discharge setting of FFT can also prove challenging, especially in A&E, as patients may be keen to go home or may be in a state where they do not want to be giving feedback. Finally, FFT is administered in the main by nurses who may have multiple

[11] In fact, in the data publication some sites achieved response rates of over 100%. This occurs when responses relating to discharges in one month are received by organisations too late for that month's submission and are submitted as part of the return in the following month. Therefore, in any given month, it is possible for the number of responses to exceed the number of people eligible to respond for that month.

[12] http://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2013/07/FFT_commentary-AE-and-IP-JAN-2014.pdf.

[13] http://www.cqc.org.uk/public/reports-surveys-and-reviews/surveys/inpatient-survey-2013.

[14] http://www.nhssurveys.org/Filestore/AE2012/benchmarks/AE12_RCB.pdf.

conflicting tasks and who may not have received training in administering a questionnaire.

4.1.2 Effect on data uncertainty (confidence intervals)

In an ideal world, FFT would collect the feedback of all patients, as this would give the "true" picture of how each site is performing. In reality, however, in the vast majority of cases only a fraction of patients respond. In these cases we cannot know for certain what the "true" profile of responses to FFT is (the response profile we would have observed had all eligible patients responded to the question). However, we can be fairly certain[15] that each "true" value lies within a certain range of the achieved result, which is known as the "confidence interval." The smaller a response rate is, the more likely the achieved value is to differ from the true value, hence the greater the uncertainty about the true result and the wider the confidence interval. Wide confidence intervals mean that large differences between the results of different hospitals need to be observed before we can confidently say that the difference is real. Likewise, changes over time of one hospital's results need to be large if we are to say with confidence that the hospital's performance has definitely got better or worse. In other words, wide confidence intervals mean that the data are less precise and hence less useful for comparing different results.

FFT data are vulnerable to wide confidence intervals because the achieved response rates can be relatively low. Also, at ward level, and in some cases at site level, confidence intervals can be wide even when the response rate is relatively high, simply because the sheer number of responses achieved can be relatively low.[16]

The size of confidence intervals means that, on average, we can only be confident that differences between FFT scores are real when the difference is over 13 points at inpatient site level or over 14 points at A&E site level (either between sites, or at the same site over time). However, for many sites the difference required will be much larger, since site-level confidence intervals vary from 0 to 103 points for inpatient sites, and from 0 to 102 points for A&E sites. At inpatient ward level, confidence intervals are even greater: on average a difference in FFT score of 30 points is required for the score differences between wards to be statistically significant. But again, considering wards individually, in some cases a much larger difference will be required, since ward-level confidence intervals range from 0 to 200 points. In other words, for some wards the confidence interval covers the entire scale of the FFT score, meaning that for these wards no possible change in FFT score can be said to be statistically significant. Only at inpatient trust level are the mean confidence intervals small enough to be able to make relatively fine comparisons (at roughly plus or minus 3 points of the FFT score on average, albeit ranging up to plus or minus 11 points at one trust). It is important to note, however, that these confidence intervals vary massively between trusts, sites and wards.

[15] All confidence intervals in this review have been calculated at the 95% level.
[16] Although the confidence interval will by definition be zero when the response rate is 100%, since the data will then comprise the views of all eligible respondents, and as such will give the "true" score.

4.1.3 Non-response effects

When not all eligible individuals complete a questionnaire, there is a further risk that the achieved results may differ from the "true" result, since the responses of those who complete the questionnaire may be systematically different from the potential responses of those who do not. This would introduce a systematic bias into the results, known as "non-response" bias.

While non-response bias is difficult to diagnose and quantify, it is likely that FFT is susceptible to non-response bias, particularly in the A&E setting. This is, in part, because FFT achieves relatively low response rates, but more specifically because it is generally offered only once to each individual, and because it is sometimes offered at a time when individuals do not wish to be detained, especially when being discharged from A&E. For these reasons, those who are less motivated to complete the test are not likely to be persuaded to do so, which means that a potentially distinct subset of responses are not collected.[17]

Non-response appears a particular issue for FFT because the nature of the bias is likely to vary from trust to trust. In centrally-controlled surveys, data collection techniques will be applied uniformly across the sample, hence non-response bias should not vary from area to area.[18] In FFT, by contrast, different trusts collect data in different ways, such as using different data collection modes and employing different techniques to encourage participation in FFT and hence to increase the response rate (see the section on 'data collection techniques,' below). For this reason, non-response bias in FFT may have a negative effect on the comparability of the data.

One challenge to diagnosing and compensating for non-response in FFT is that we do not know whether we are achieving lower response rates with particular demographic groups, since demographic data of this sort is not routinely collected (see the section on 'Data adjustment,' below). It should also be noted that non-response effects may also arise out of selection bias (see the section on 'Other biases,' below).

4.2 Data collection techniques

In FFT, 'data collection techniques' refer to the methods by which the FFT question is asked of patients and their responses are recorded. FFT is not centrally controlled and currently employs a permissive data collection methodology, meaning that, while the timing of the test is specified as being at or within 48 hours of discharge, sites are free to collect the data by any suitable means.

As discussed in the introduction to this section, since the implementation of FFT, trusts, and different wards within trusts, have found a range of solutions for conducting FFT and for meeting the CQUIN response rate. The diversity of hospital

[17] Cf. Groves et al. (2009) *Survey Methodology*, Second edition. Hoboken, New Jersey: John Wiley & Sons: pp. 196-201.
[18] Although demographic differences between areas may lead to differences in non-response effect.

management approaches, architectures, patient types and existing patient experience research programmes mean that different approaches for FFT may work better in different settings. This empowerment with regard to designing the implementation of the FFT may help explain the strong sense of local ownership of the test that was observed by the qualitative review, and the encouraging signs of the data being digested and acted upon by staff. One criticism of centrally-controlled, standardised patient experience surveys has been that their results have not always been understood and acted on by staff. [19] By contrast, the FFT decision to permit a variety of approaches appears to have been successful in achieving "buy in" from staff.

However, the employment of different approaches to data collection has implications for the comparability of data. In this section we will investigate how the data collection techniques used in FFT affect the conclusions we can draw from the data, and what the limitations are.

4.2.1 Data collection modes

The data collection 'mode' refers to the overall method by which data is collected. In social and market research surveys, the most common modes of data collection are face-to-face interviewing, telephone interviewing, postal questionnaires and online questionnaires.

In FFT, following the permissive methodology outlined in the Department of Health guidance, a number of different data collection modes are used between sites, and many sites or trusts employ a number of different modes of data collection. Multiple data collection modes within a single setting can be used either to better meet the response rate target, or to better represent different voices (since some groups may be more likely to complete the test by one mode than by another).

Postcard questionnaires completed within the hospital represent the majority mode of data collection in both the inpatient and A&E settings (at 80% and 49% respectively).[20] Half of inpatient trusts use a single mode of data collection (50%), and one third of A&E sites use a single mode (32%).

4.2.2 Mode effects

Survey methodological literature shows that different data collection modes can result in different results being obtained.[21] This can be due to a change in the effect of existing biases as a result of the mode: in particular, non-response bias tends to vary with mode (e.g. some individuals are less likely to respond by pen and paper, while others are less likely to complete the test online). In addition, the mode effect

[19] Robert, G. and Cornwell, J. (2013) 'Rethinking policy approaches to measuring and improving patient experience,' in *Journal of Health Services Research & Policy* 18(2): 67–69.
[20] This is the proportion of the total volume of national responses for January 2014. Data available at: http://www.england.nhs.uk/statistics/statistical-work-areas/friends-and-family-test/friends-and-family-test-data/.
[21] Groves et al. (2009) *Survey Methodology*, Second edition. Hoboken, New Jersey: John Wiley & Sons: pp.163-173.

can arise from biases that are specifically introduced by a particular mode: for example, responses to interviewer-administered modes (such as telephone) tend to be more positive than responses to self-completion modes (e.g. pen and paper), or tend more often to correspond to the responses that the interviewee expects the interviewer to favour (an example of social desirability bias).

In the FFT setting, mode effects are difficult to isolate because there are other research biases across sites (and hence a lack of control measures). Nevertheless, quantitative analysis for this review, which investigated nationally-collected data along with data provided by two FFT suppliers, suggests that mode effects may significantly affect FFT data.[22] In particular, online responses appear to elicit lower FFT scores than do paper responses. These findings are corroborated by recent analytical work by the Picker Institute.[23]

4.2.3 Timing of the question

One significant contributing factor to differences between modes may be the differences in timing between certain modes, and the corresponding difference in setting when completing the FFT. Whilst some gather patient opinion just as the individual is about to return home following treatment, other modes gather responses at any time within 48 hours of discharge. Opinions are therefore given in very different contexts, with point-of-discharge judgements being particularly vulnerable to the "halo effect" (a form of gratitude bias), while judgements provided later in the 48 hour window may benefit from greater reflection on the care experience. The quantitative review found statistically significant differences between sites that predominantly used point-of-discharge responses and sites that collected at least 10% of their responses after discharge. In both the inpatient and A&E settings, post-discharge data collection was found to result in lower FFT scores.[24]

4.2.4 Other biases

Survey methodology literature suggests that differences in how data collection modes are implemented are likely to affect responses to questions. FFT data may be subject to a range of biases to varying degrees, over and above biases associated with the data collection mode and the timing of the question. (Note that data from hospitals using FFT suppliers are equally vulnerable to these biases, not least because hospitals using suppliers often still use their own front-line staff to collect the data, with the supplier providing only a data processing and analysis service.) The biases arise from the following differences in approach:

There are differences, for any given mode, in how the question is administered. Social desirability bias may be an issue depending on whether a staff member distributes the question, whether the respondent is given privacy when completing FFT, or whether the respondent requires assistance in completing it.

[22] NHS England (2014) *FFT Review: Quantitative strand*: pp.17-20.
[23] Picker Institute Europe (2014) 'Demographic and methodological variation in the NHS Friends and Family Test: secondary analysis of existing data.' Manuscript submitted for peer review.
[24] NHS England (2014) *FFT Review: Quantitative strand*: pp.18-19.

The priming effect may be relevant, depending on how, if at all, the question is presented to the respondent: whether the purpose of FFT is explained, or whether the scoring system is explained.

There are differences, for any given mode, in how the materials are designed or presented (e.g. layout or branding of postcards, introductory screens or layout of webpages). This can result in the framing effect.

Different measures are being used to ensure all are offered the question; there is also evidence of staff sometimes actively choosing not to offer FFT for non-clinical reasons. As such, data are subject to selection bias.

Different measures are being used to maximise response rates, such that non-response bias may vary across sites. Differences in the number and/or nature of questions being asked in addition to the FFT may also impact on non-response bias.

Given the setting of the FFT, clinical discretion is sometimes required with regard to the appropriateness of asking the question. In cases where a patient is distressed or is less able to participate (for example, when the patient has a learning difficulty or has dementia),[25] it may be deemed inappropriate to offer the FFT. A certain amount of subjectivity may be involved in such decisions, hence there may be some variation between sites in the frequency of FFT not being offered for clinical reasons.

4.3 Demographic differences

Different NHS trusts provide different ranges of treatments in different volumes, serving populations of different demographic compositions. Evidence from research literature shows that differences in survey response are commonly observed between different demographic groups.

Currently there is no mandate for demographic data to be collected as part of the FFT. This supports the principle that FFT should be as simple as possible, with minimal burden on the respondent, which in turn helps to maximise response rates. However, some trusts routinely request demographic data as part of the FFT, often as part of a standard supplier offer.

Analyses of FFT data from two such suppliers (Picker and iWantGreatCare) suggest significant differences in FFT score between different demographic groups. As such, trusts serving populations with different demographic profiles may expect to achieve different FFT scores for a similar level of care.

4.4 Comparability of data

While the qualitative review has provided strong evidence of local use of FFT data for service improvement, the original ambitions for FFT also included using it to compare the relative performance of providers of NHS services. In this section we review the

[25] The latter situation will be mitigated in future through the 'Access for All' scheme, through which FFT materials that are more suitable for these audiences will be disseminated.

implications of the quantitative review of FFT data for comparing data between trusts, and different options for FFT going forward.

Comparison of patient experience data between trusts should where possible take into account the nature of the patient population served by the trusts. Demographic differences have been shown to result in different responses to the FFT. In addition to this standard principle of comparison, however, FFT presents several other challenges for data comparison.

4.4.1 Current comparability

4.4.1.1 Biases

Given the evidence presented in the sections above, FFT data should not be considered statistically comparable at an inter-trust level. In particular, recorded mode effects imply that it may be invalid to directly compare FFT results that have been collected using different modes – a problem that is further confounded by the fact that a significant proportion of trusts employ multiple collection modes. One particular aspect of this issue appears to relate to the fact that different collection modes gather responses at different times: either at the point of discharge (postcard, kiosk, tablet, token) or at some point after discharge (posted questionnaire, online, SMS, telephone). In addition, a number of other biases further compromise the comparability of FFT data across sites, which result, in the main, from the fact that no standardised form of data collection materials are used within any particular collection mode, and, more crucially, from the fact that front-line staff tend to collect the data themselves.

4.4.1.2 Statistical uncertainty

Where a site's approach to FFT data collection remains the same over time, however, such that biases inherent in the data collection approach remain reasonably constant, it would be valid to compare that site's FFT results over time. This type of longitudinal comparison should nevertheless be practiced with caution, since the large confidence intervals associated with the data mean that differences in FFT scores at inpatient site and ward levels, and A&E site and trust levels, are generally not statistically significant unless the difference is relatively large. In many cases this is also true at inpatient trust level. This statistical uncertainty also applies to data between sites, such that, even if a fully standardised data collection were implemented, rankings of sites or wards by FFT score would be misleading unless response rates were greatly improved.

4.4.2 Future options

4.4.2.1 Data adjustment

As discussed above, FFT data is subject to a number of identified and probable biases. In some situations survey biases can be compensated for by using data adjustment techniques.

One clear source of non-response bias occurs when patients of particular characteristics are underrepresented in the data: as mentioned above, demographic differences have been observed in FFT responses. In many surveys a "case-mix adjustment" is routinely made,[26] in order to ensure that the final results are representative of the demographic profile of the population being investigated. In these cases different weights would be applied to the responses from people of different genders, ethnic origins, age bands or socio-economic groups. Such an adjustment is not currently possible for FFT, however, as the demographic profile of respondents is not currently standardly recorded. Were it to be then there would be a greater potential to analyse the data and draw greater insight from it particularly related to relative experiences across different groups. The future collection of demographic data via FFT would provide one way of mitigating against non-response bias, along with ensuring that the views of patients with protected characteristics are fairly represented.[27]

Likewise, some mixed-mode surveys employ post-fieldwork adjustment factors to negate identified differences resulting from mode effects. Under controlled conditions it would be possible to measure the difference in responses between modes of FFT and to adjust the data accordingly, but such controlled conditions cannot currently be obtained in the context of FFT, due to the number of differences between how each mode is implemented at the local level. If it were decided that case-mix adjustment or mode adjustment of data may prove beneficial for FFT, then controlled experiments could be conducted in order to better understand the feasibility of these approaches.

4.4.2.2 Standardisation of the methodology

One option to improve the data quality of FFT would be to mandate a single mode of data collection. A different, slightly less restrictive option would be to standardise the timing of the FFT question. In the inpatient setting it would perhaps be most appropriate to limit collection to the point of discharge, since this appears to increase staff engagement with the data. In A&E, however, post-discharge collection could be more feasible as a single option, given the pressures experienced by A&E staff, the more limited opportunity to administer the FFT in this setting (given the brevity of care episodes), and existing difficulties in meeting response rate targets. Standardising the timing of the FFT question for each setting would ensure that "what" is being measured has a greater degree of uniformity across all sites. However, due to the permissive nature of local administration of FFT, neither of these options would eliminate all biases associated with the data. Moreover, the cost on trusts needing to change collection modes could be great and could impact on wider existing patient experience programmes that are being used successfully. The removal of the chance to collect data after discharge could also make it more difficult for trusts to meet the response rate requirements, particularly in the A&E setting.

[26] See, for example, this note on weighting in the national GP Patients Survey: http://www.gp-patient.co.uk/faq/weighting/.

[27] It should be noted, however, that any form of lengthening of a questionnaire can negatively impact on response rates, because the process of completion becomes more burdensome. We may also expect to see some item non-response for demographic questions.

In order to produce fully comparable FFT data, such that it would be suitable for performance management purposes, with financials attached, a radical shift in the organisation and administration of the test would be required. In order to eliminate substantial differences between FFT approaches across trusts, the test would need to be administered centrally, with a third party distributing the test via a single data collection mode, and collecting responses directly from patients, rather than the trust having a role in the collection. Demographic data would be recorded along with the FFT response and compared with a database of eligible patients, so that data could be adjusted to match the demographic profile of the eligible population and/or to normalise the demographic profile of trust scores. In other words, the FFT would need to employ a methodology akin to a nationally representative survey, such as the Inpatient Survey and the GP Patient Survey. The trade-off would be that FFT data would not be available in real time. The cost implications of a centrally administered FFT, as compared to the current approach, would also need to be investigated.

Given that a centrally administered census approach to FFT might not be cost-effective, a sample-based approach might perhaps be considered. However, since hospital attendance cannot be predicted, these samples would need to be supplied retrospectively, meaning that FFT could no longer be asked at the point of discharge. A sample approach would also mean dispensing with a fundamental principle of FFT: the opportunity for all service users to feed back.

Achieving data comparability by moving to a centrally-administered model could strip FFT of many of its most distinctive characteristics, including local ownership and real time feedback:

- FFT data would no longer be available to staff in “real time” (i.e. the FFT feedback would not be available to frontline staff very soon after being collected):
    - Since data would be collected by the supplier, staff would not be able to immediately view comments;
    - Since data adjustment processes would be required in order to ensure comparability, staff would not be able to view their results for a significant period of time (given the size and complexity of the FFT collection, and the need for central coordination, this could be a matter of months, not weeks).

These risks are particularly salient when considered in light of the underuse of nationally collected survey data by trusts. This problem was recently summarised by Glenn Robert and Jocelyn Cornwell in the *Journal of Health Services Research & Policy*:

> *“[L]eaders of local health care organizations have largely not used the results [of national surveys] to formulate their own strategic goals to improve patients’ experiences. […] It appears that the very existence of national surveys has contributed to a tick box or compliance mentality on the part of management boards. The survey programme may have lulled them into thinking that they were paying attention to their patients’ experiences. For instance, and despite examples of good practice, a review of management*

board agendas and minutes from a sample of hospitals found that patient experience data were rarely used to spark debate and action."[28]

The qualitative FFT research by Ipsos MORI found numerous examples of staff viewing very recent feedback and being able to clearly understand any problems it raises. This suggests that this "proximity" to the feedback, in terms of being able to associate it with specific recent incidents and experiences, is an important driver of FFT's use for service improvement.

4.4.2.3 Removal of small numbers from the presentation

Wide confidence intervals present a further problem when hoping to compare data, including when comparing data from a single site over time. Analysis has shown that this problem cannot be simply resolved by only considering sites with large numbers of responses: for example, if we excluded inpatient sites with fewer than 100 responses per month, then we would be left with mean confidence intervals of 7 points – a reasonable result – but at the cost of excluding 47% of sites from the publications.

Aggregating data over several months could provide a more inclusive solution. By collating six months of data, mean confidence intervals would reduce to 6 points for inpatient sites and 5 points for A&E sites. However, this would be at the cost of dampening any significant changes in score over the course of those six months, making it more difficult to spot important changes over time. Even more significantly, perhaps, the "real-time" nature of FFT would be compromised (although data could be privately viewed by relevant bodies in real time, with publication of the collated data coming later). Any decision on suppressing or collating data would also need to consider the FFT's commitment to transparency on the one hand, and its duty of confidentiality towards its respondents on the other hand. These issues are discussed in section 6, below.

[28] Glen Robert and Jocelyn Cornwall (2013) 'Rethinking policy approaches to measuring and improving patient experience,' in *Journal of Health Services Research & Policy* 18(2): 67–69.

5 The FFT score

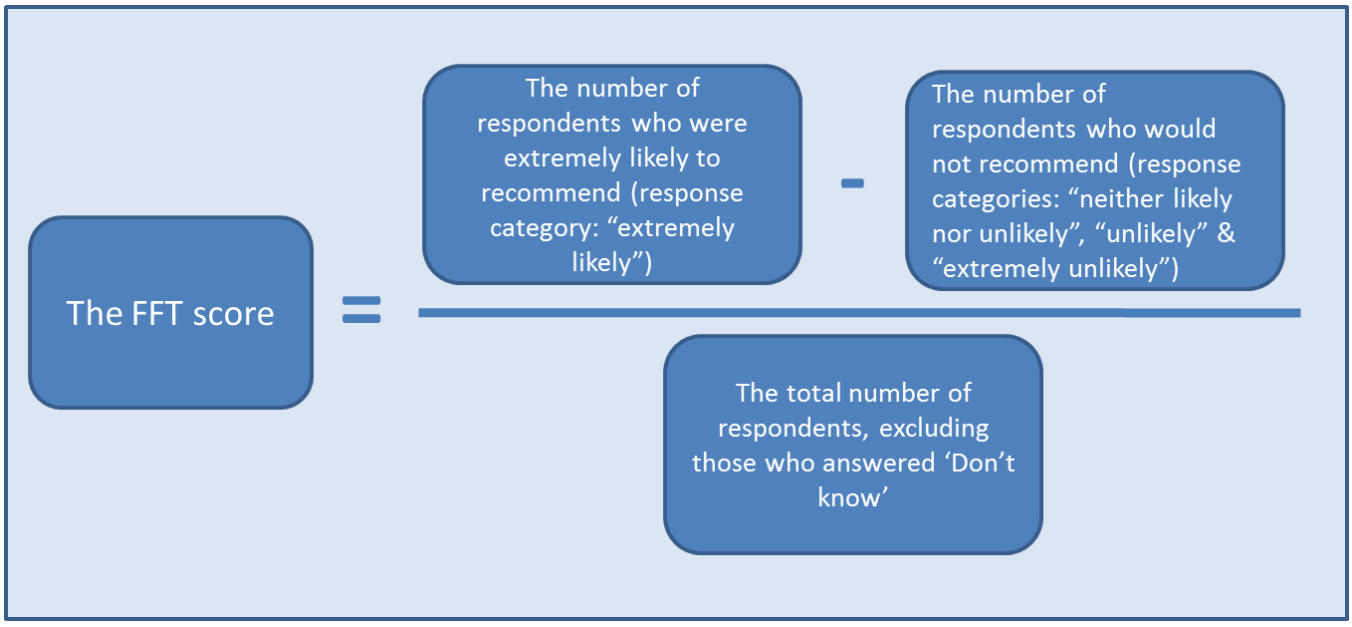
The FFT question, response scale and scoring metric were all inspired by the private sector Net Promoter Score (NPS), a customer loyalty tool.[29] In this section we discuss the performance of the FFT scoring system (known as the "scoring metric") against the original ambitions set out for it. We also consider alternative scoring metrics, and assess their suitability for FFT.

5.1 Purposes of the score

The FFT collects six types of response, yet there is a need to summarise FFT data in a single "FFT score." The score is produced via a "scoring metric," which combines the data from the response scale in a defined way. An FFT score should help to support the patient choice and service improvement functions of FFT by providing a simplified way for the public and frontline staff to judge relative performance in the FFT. In order to avoid confusion, and to promote transparency, a single score is used for all audiences.

When a set of responses is transposed into a single score, certain features of the information are inevitably lost. Nevertheless, it is desired that the FFT score provides a reasonable representation of overall performance. In particular, the FFT score should reward positive performance, as well as being responsive to negative feedback. As such, the FFT score should ideally discriminate between different levels of performance, and in particular should act as an early warning system if a site is producing concerning results.

The current FFT scoring metric is defined as follows:



[29] A registered trademark of Fred Reichheld, Bain & Company and Satmetrix, NPS was introduced by Fred Reichheld in a 2003 journal article: Reichheld, Frederick F. (December 2003). 'One Number You Need to Grow".' Harvard Business Review. For further details, see http://netpromotersystem.com/about/index.aspx or http://www.netpromoter.com/why-net-promoter/about-net-promoter/.

5.2 How the score is being used

This section examines how different organisations in the health system interact with the FFT score, how they use it, and what obstacles exist to its use as a single, uniformly understood measure. The FFT score is currently calculated at ward, site and trust level, and is aggregated nationally each month (both separately for the inpatient and A&E settings, and in a combined form for trusts).

5.2.1 Frontline staff

Qualitative research with ward staff found that the score is not widely understood. This is partly due to the way that it is calculated, with many finding the figure too abstract a representation of the feedback: staff find it difficult to judge whether any particular score is good or poor. The score is often wrongly understood and communicated as a percentage, which can add to the confusion. In summary, staff do not find the score to be intuitive, and that this presents an obstacle to its wider use.

Many trusts have decided against using the official FFT score, opting for different ways of presenting the data at a local level. These include using the 'overall recommend' figure (expressed as a percentage), and showing the profile of all 6 response options. Where suppliers are used for the processing of FFT data, the monthly reports provided to trusts sometimes use alternative presentations of the data. One supplier to a large number of trusts provides ward-level reporting that excludes the actual ward FFT score. Instead a star system is used, which translates each FFT responses to a star rating of between one and five and then presents the average. This tends to give a more positive picture than the FFT score. In trusts that use this supplier, awareness and understanding of the official FFT score tended to be lower amongst frontline staff.

A further, widespread criticism of the FFT score among staff is that the scoring metric is unfair. In particular, many perceive that responses of 'likely to recommend' "do not count" in the current system. This is considered unjust by many staff, since 'likely to recommend' is felt to be a good, positive response. In addition, many question the legitimacy of counting the neutral 'neither/nor' response as a detractor. Aside from damaging the reputation of the FFT score amongst staff, the qualitative review found evidence that these aspects of the current system may be motivating "gaming" amongst staff. In a number of cases staff reported that they explain the scoring system to patients ahead of giving them the FFT, so that they know that 'likely to recommend' does not count. This "coaching" of patients may have a direct influence on the way they choose to respond, thus skewing the FFT results on the ward and meaning that the data does not give us a true picture of patient experience. Also, the fact that such coaching may not happen on the majority of wards, while on others patient responses are being influenced to a greater or lesser degree, means that the comparability of FFT responses between wards and trusts is compromised.

5.2.2 Boards of trusts

Board-level staff's main exposure to the FFT score was found to be through monthly board reports, where FFT data was accompanied by other patient experience data on a summary dashboard. Some board members mentioned similar issues about the scoring to those of frontline staff, in particular that the measure was not intuitive, was difficult to evaluate, and was easily confused with a percentage. In general, however, Board staff found the score less problematic than did ward staff, and understood the score as an important metric for monitoring the performance variation across their trust. In some cases, board papers included a caveat advising caution when interpreting the score.

5.2.3 By central bodies

In the qualitative research with stakeholders from organisations such as CQC, Monitor, the Royal College of Nursing, the Foundation Trust Network and the Department of Health, critical comments about FFT centred on the score, which was regarded as a relatively insensitive measure – one which doesn't convey enough information in and of itself. Linked to this was the feeling that the addition of a follow-up question (establishing why the respondent said what they did) was a crucial element to understanding what improvements people were calling for.

5.2.4 By the Public

As noted in section 2, above, in the qualitative discussion groups, patients and the public expressed dissatisfaction with the current FFT score. In particular, it was found to be difficult to understand and interpret and was easily confused with a percentage. The participants expressed a clear appetite for an overall score that is simple and easy to understand, but many also expressed the desire to have access to as much information as possible. Participants wanted to be able to explore the full set of FFT responses, because different individuals are interested in different aspects of the feedback: while some are most concerned about the number who would recommend, others are keen to understand the number of poor responses a site has received. Participants also wished to know the number of responses the feedback was based on, so that they could make a judgement about the reliability of the overall score: this is a practice some were accustomed to from viewing online reviews about products or services.

5.3 Alternative scoring metrics

In light of the criticisms of the current scoring system outlined above, this section considers the suitability of alternative scoring metrics for FFT. The options considered in the quantitative analysis and the qualitative research were selected based on the findings of an earlier review of FFT scoring options,[30] on alternative

[30] Ipsos MORI (2012) *Scoring and presenting the Friends and Family Test: A review of options*. Research conducted on behalf of the Department of Health and NHS Midlands and East. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/214942/FFT-Ipsos-Mori-research-report.pdf.

presentations currently in use on some wards and by some suppliers, and on presentations of user scores on popular websites:

- A “star system,” which converts each FFT response to a star rating between 1 and 5, and which presents the average star rating for the ward, site or trust (to one decimal place).
    - This system is used by some FFT suppliers, and resembles rating presentations used on popular websites, such as Amazon and Trip Advisor.

- A weighted score out of 100, which converts each FFT response to a score of 0, 25, 50, 75 or 100, and presents the average.
    - This system is an alternative way of presenting a score to the star system.

- A “Net positive” score, which presents the proportion of positive responses *minus* the proportion of negative responses (discounting ‘Don’t Knows’).
    - This follows a similar logic to the current FFT score (a net score), but counts all “likely” responses as positive, and does not count neutral responses as negative.

- The percentage of people who are likely to recommend the service.
    - This is the simplest option, in that it is a simple sum of positive responses, both ‘Likely’ and ‘Extremely likely,’ and does not account for negative responses.

- A positive *and* negative scoring system, which presents both the number who are likely to recommend *and* the number who are unlikely to do so.
    - This does not present a single score, but responds to evidence that the public wish to view both of these figures. Given the concerns about the direct comparability of data between sites, the use of a double metric that makes comparison more difficult could be viewed as advantageous, being potentially less likely to mislead the public.

For the quantitative analysis, responses from three consecutive months of FFT (Sep – Nov 2013) were inputted into the different scoring metrics, and the resultant scores were assessed in three different ways:

- Validity testing, to measure the extent to which the metric captures variation in numbers of positive and negative responses to the FFT;

- Discrimination testing, to measure the degree to which the metric differentiates between different sets of responses (and hence creates unique scores for different sites);

- Stability testing, to measure the degree of fluctuation in site scores and rankings from month to month.

While no single scoring metric outperformed all others in all tests, the analysis produced a number of interesting findings:

- Since responses to FFT are so heavily skewed towards the positive, the current scoring system achieves the widest variation in scores. This is because it does not count "likely to recommend" as a positive response, and because it negatively counts neutral responses: in other words, variation is achieved at the cost of imposing a non-intuitive weighting system on the responses. All other options present little variation in the data, especially the star system, the weighted score out of 100 and the total positive responses.

- Unsurprisingly, given its treatment of "likely" and neutral responses, the current metric proved least adept at capturing variations in numbers of positive and negative responses to the FFT. Ultimately, this aspect of the metric is an evaluative problem: we must decide what should count as a good response and what should count as a poor response. However, it appears the most effective way of capturing the overall character of FFT feedback is to present two headline metrics rather than one, showing both the total who would recommend and the total who would not.

- The current metric produces the greatest number of unique scores for sites (and hence the greatest number of unique rankings), and the star system produces the fewest, at just 8 unique rankings (all between 4.3 and 5 stars out of 5). However, given that confidence intervals around all scores are large, it could be perceived as advantageous to not present fine-grained differences that are less likely to be statistically significant.

- Of the options considered, the star system and the current metric produced the greatest stability in the relative performance of sites over time, with the Net Positive score and the total positive responses producing slightly more volatility.

Qualitative research with the public had much clearer outcomes with regard to the comprehension and salience of the different scoring metrics:

- The net positive score and the weighted score out of 100 were thought to be too difficult to understand and interpret. In particular, the latter was confused with the percentage of people who recommended the service

- The star system was intuitively understood since people were familiar with its use on websites. However, some questioned whether such a system, generally associated with consumer services and products, was appropriate for what they saw as the more serious context of healthcare. It should also be noted that, unlike ratings on Amazon and Trip Advisor, the scores on FFT will be universally clustered around 4.5 or 5 stars, and that individual reviews with comments will not be available, such that this presentation may be less effective at informing choice in this context.

- The percentage of people who recommended was viewed by some as the simplest possible measure, but for some individuals the percentage of negative scores was more interesting and important.

- In general, there was an appetite to be able to drill down into the FFT results, looking not only at the full set of responses for a hospital, but also the number of responses it was based on, and a sense of how the hospital's results compare to others.

It should also be noted that, while the qualitative groups found interest in FFT data, the participants wished to view FFT scores alongside other relevant data on hospital quality, such as staffing rates, infection levels and waiting times. They felt that, in situations where they have a choice about the NHS service they will use, FFT data would form just one of several factors that would inform their decision.

5.4 Implications of changes to the score

Altering the headline FFT score would incur a relatively small burden on hospitals and suppliers when compared with other possible changes. The change would simply affect internal reports that present the current FFT score, and modifications to these should be simple to make.

Since any new scoring metric would use the same raw materials as the current FFT score (i.e. the same question and response scale), historical FFT data could be used to produce trend data for the new scoring metric.

5.5 Other possible modifications

5.5.1 Changes to the question

Qualitative research encountered mixed views about the current FFT question from staff, stakeholders, patients and the public. Most patient and public group participants quickly understood the principle of the question, feeling it would be easy for them to answer, and many felt that the notion of recommending a service to friends or family members emphasised that a certain level of importance and excellence was in question. Nevertheless, criticisms of the FFT question were encountered across the research: concern about the use of "brand loyalty" concepts in the health sector; inappropriateness of recommendations when discussing treatments of poor health; perceived lack of choice in many health settings (such as A&E) and for some audiences, thus making the notion of recommendation somewhat incoherent.

One cost of changing the FFT question would be the loss of existing trend data. More significantly, existing data collection materials, such as printed cards and software, would need to be altered in light of the change. The cost to trusts is unknown, but significant advance warning of the change would be appropriate in order to allow trusts to make plans to minimise the cost of change.

The present report does not explore alternative FFT questions because, in its review of the first six months of FFT, strong evidence supporting a change of question was not found, with other possible modifications found to be of higher priority. If clear reasoning and support for a change of question was found, then alternative questions would need to be tested for validity and reliability.

5.5.2 Changes to the question response scale

The qualitative review also uncovered some criticism of the current FFT response scale, particularly with regard to the most positive response, “extremely likely to recommend.” Many participants in the public groups found this response wording too effusive for it to be a practical option for people of British culture, even when the service received was excellent. However, it should be noted that “extremely likely to recommend” is by far the most frequently selected response to the FFT, and the response distribution that FFT exhibits is strongly skewed towards the positive.

The length of the response scale was also discussed in the qualitative review, with views divided. Some, mostly clinicians, argued for a longer, 11-point scale (from 0-10) in the hope that this would introduce greater differentiation into the data than is produced by the current scale. Some members of the public, however, thought the current system was overly complicated and that a three point scale (positive, neutral and negative) or even a binary (would/would not recommend) response scale would be preferable.

It should be noted that altering the response scale would come at the cost of potentially losing all trend data, since responses to one scale generally cannot be compared with responses to a different scale. Given that FFT remains in a period of integration in the service, it may be unwise to alter the scale at this stage. Trusts are currently experimenting with ways of deriving the greatest value out of FFT, and the use of FFT to monitor changes over time is potentially one of its greatest strengths. As such, it may prove propitious for FFT to demonstrate the extent to which it can provide insight over time before a change to the response scale is investigated.

A further consideration is that a change to the response scale would require trusts and suppliers to modify data collection materials and processing techniques, so it would be important to build a lead-in time into the process to ensure that prior investments are not wasted and so that new materials can be prepared.

If an alternative scale were to be considered, then controlled experiments should be run in order to compare the response distributions achieved by the current scale and the alternative, and cognitive testing should be utilised to understand how each scale is being interpreted.

6 FFT as a local feedback mechanism

Following the description in section 2, above, of how FFT is currently being used within trusts as a service improvement tool, this section synthesises the evidence in order to present findings of best practice in the use of FFT at a local level. It also discusses the obstacles that exist to making FFT work as effectively as possible to improve patient experience.

6.1 Best practice models of data usage

The participation of staff in FFT collection was found to create a more immediate sense of ownership of the data, a greater curiosity to review it, and encouraged them to take responsibility for addressing issues raised by patients. However, we acknowledge that it may not always be appropriate or feasible, depending on the care setting, for frontline staff to directly collect the FFT data. More important is that FFT feedback is clearly communicated to staff and that they are able to act on it.

A strong finding of the review is that FFT appears much more valuable for service improvement when qualitative feedback is attached to the response to the main FFT question. This question should be an open follow-up, thus allowing the patient to make any further comment, good or bad, no matter how they responded to the main question.

The research undertaken amongst staff groups demonstrates that rapid feedback of scores and comments to ward managers and nurses results in a number of positive outcomes. The immediacy of feedback gave greater salience to the comments, both positive and negative, helping staff to better understand their patients' experience of their care. The rapid turnaround of FFT data means that problems can be identified quickly and changes made.

Given FFT's potential for highlighting issues, particularly in light of the Francis review, the best practice use of FFT involves taking an initial view of the data as soon as possible after it is submitted – daily, if the data collection solution allows it. Delays in exposure to the negative comments clearly impact on the speed with which any change or improvement is implemented, and this blunts the effectiveness of FFT as a real-time response mechanism.

All negative FFT feedback should be treated as potentially an indication of harm or risk and should be reviewed both by the ward or departmental manager and an appropriate hospital- or trust-level figure, such as the Patient Experience lead.

Collated FFT data should be discussed periodically in ward or departmental meetings, with the quantitative findings explained to staff and a selection of verbatim comments reflected on. Plans for responding to FFT feedback should also be routinely discussed. Where collated FFT data is only available monthly, via the national submission or through a bespoke report, there appears a greater burden on patient experience leads and ward managers to have effective processes in place for communicating the results in order to engage staff.

FFT should be visible to both patients and frontline staff, further embedding it in the hospital culture. Posters within the wards should communicate the nature of the responses on the ward, and if possible also provide evidence of how the feedback is being used to make a difference, thus demonstrating its value both to patients and staff.

As discussed in section 2, above, individual quantitative FFT responses can appear abstract, such that greater value is often found in the qualitative follow-up comments. As such, in order to derive value from the quantitative FFT data at a local level, there is a burden on staff to collate and interpret the data, perhaps considering performance over time, or comparing performance with that of similar wards within the same hospital.

6.2 Challenges and opportunities

The greatest challenge and opportunity for FFT lies in integrating its use within services such that all staff engage in an efficient and effective process of quality monitoring and continual striving for improvement. The best practice advice presented above may assist in the more effective use of data, but ultimately service leaders need to encourage and enable a culture of continuous patient experience data collection and learning. This will require management structures for disseminating findings and designing responses and action plans, along with analytical tools for understanding data over time and gleaning insight.

Such large-scale data collection and analysis also requires significant resource, particularly as FFT is rolled out to other services, which suggests that investment in technology may ultimately be required. Electronic data collection methods, such as the use of tablets, eliminate the burden of collating paper-based data and allow for the immediate sharing of data with staff.

However, it should also be noted that the democratic principle of FFT must also be upheld: all patients should have a fair opportunity to give feedback, and some may be less comfortable or capable with more technologically advanced formats. As such, advancements and best practice must go hand in hand with ensuring FFT is accessible for all.

Another significant challenge and opportunity for FFT concerns the use of the qualitative feedback data. This data is arguably the richest part of the FFT feedback system, but collecting and processing large volumes of qualitative data from the free text question presents a significant challenge to trusts. Even in trusts where the qualitative data was being reviewed and acted upon, this tended to be done on a fairly basic level, with Patient Experience Leads scanning printed pages to identify negative comments, and then using these to build ward level action plans. A thorough and consistent approach to categorising the free text comments is currently beyond the capability of most trusts. Coding is time consuming, and requires specific skills that are not necessarily to be found within Patient Experience teams, which are also likely to be stretched for resource.

Electronic methods greatly reduce the barriers to analysing qualitative data, not least eliminating the need to scan or input this data. Software that analyses qualitative

feedback for theme and sentiment offers the possibility of enabling greater learning from comments, but the effectiveness and accuracy of this software is not yet self-evident, hence this would need to be investigated and established before a roll-out of such an approach could be recommended. Nevertheless, the opportunity afforded by this data is significant.

However, such methods should not be used at the cost of distancing frontline staff from the data: arguably FFT is at its most powerful when a member of staff reads a comment on a postcard at the end of a shift and is made aware of an example of excellent care, or of a problem that requires urgent attention. Moreover, the fact that a patient has taken the time to provide qualitative feedback means we have a duty to ensure that their voice is heard and accounted for in some way, ideally by having at least one person read each comment in full, rather than relying on software to understand and summarise it.

This challenge and opportunity with regard to qualitative data concerns various feedback mechanisms and various organisations across the health sector. The sector is experiencing a growing accumulation of qualitative patient feedback via media as diverse as FFT, Care Connect, NHS Choices, complaints pages, Patient Online and Twitter. This data potentially houses great insight for NHS England, the Department of Health, the Care Quality Commission, and other organisations concerned with patient experience, care quality and patient safety. Designing an organised and coordinated approach to reviewing and analysing this data, such that every patient's feedback is counted and learned from, should be a priority for the sector.

7 Publication of FFT data

In this section we will review how the FFT data has been published thus far, from the national publication to its presentation in hospital wards and A&E departments. Drawing on the findings of the qualitative research and quantitative data analysis, we will also consider the future options for the publication of FFT.

7.1 National publication

Central to the philosophy of the FFT is the publication of its data for ensuring transparency. Each month trusts submit their responses to NHS England, along with tallies by data collection mode, and NHS England then publishes the full dataset, including FFT scores. This FFT data is presented at trust, site and ward level for the inpatient collection, and at trust and site level for the A&E collection. Also, irrespective of the FFT's transparency agenda, NHS England is committed to publishing centrally-collected FFT data by virtue of FFT's designation as an experimental official statistic.

7.2 Data suppression

In order to protect the anonymity of respondents, NHS England currently suppresses FFT data where there are small numbers of responses for the unit being reported, such as the ward or the site. Anonymity in the context of FFT is thought to be particularly important, because some respondents may be concerned that negative feedback could impact on the care they receive in future. NHS England's current approach is to suppress the FFT response profile (the tallies of individual FFT responses, i.e. 'extremely likely,' 'likely,' etc.) for any unit where the total number of responses for the month totals fewer than five. This is because it is often possible to deduce individual responses from the FFT score when low numbers of responses are involved. The number of responses and the overall FFT score are not currently suppressed in the presentation of data.

In a typical month around 10% of inpatient ward-level responses and around 5% of inpatient site-level responses are suppressed.[31] Site-level suppression often occurs as a secondary or tertiary suppression, which sometimes has to be performed to try to ensure that a suppressed figure cannot be identified by deduction from other figures.[32] This approach to suppression is complex to apply and does not always achieve its goal of concealing small numbers of responses. In some cases, suppressed responses can still be worked out from the score and the number of responses. As such, NHS England would like to alter the rules for suppression, since the current system is not achieving its stated purpose, of successfully concealing individual responses when numbers are low, while on the other hand not being as transparent as possible. In addition, the current flawed system requires significant analytical resource.

[31] NHS England (2014) *Friends and Family Test Suppressions: A review*.
[32] For example, in a site comprising two wards, where the data from the first ward requires suppression due to fewer than five responses, then the responses of the second ward would also require suppression, since otherwise the responses of the first ward could be calculated by subtracting the responses of the second ward from the total responses for the site.

Alternative approaches to suppression may include one or more of the following strategies:

- Publishing ward level data on a quarterly basis only;
- Not publishing data at ward level;
- Increasing the number of unit responses for which suppression takes place;
- Suppressing the FFT score and/or the number of responses in addition to the individual response tallies for units with low numbers of responses.

These options reflect a number of considerations. Raising the number of responses for which suppression takes place and/or suppressing the score in addition to the response profile should make the suppression more difficult to unpick. It would also reduce the burden and complexity of the monthly suppression, as would the strategy of publishing ward level data less frequently, or not publishing ward level data at all.

Qualitative research with patients and the public suggests that ward level data is not especially salient to them, given that few patients have the opportunity to choose the ward they stay on. Likewise, the removal of ward level data would improve the overall quality of the published data, given the very large statistical uncertainty associated with much of the data at this level. However, NHS England would strongly encourage trusts to continue to collate their data at ward level, not only to support and encourage the use of FFT for local service improvement, but also to create a competitive atmosphere between wards if desired.

Another option to be considered is that of publishing all data, irrespective of small numbers. This solution would necessitate making clear to patients exactly how the data would be processed and published, and the potential effect of this on the response rate and the nature of responses should be investigated.

However, given that raw FFT feedback is often available to be consulted by frontline staff (e.g. in some cases staff can look at the FFT postcard responses), we should note that the FFT is currently unable to guarantee anonymity to its respondents, irrespective of the policy on data suppression at the national level. It is reasonable to assume that immediate identification by frontline staff may be of greater concern to patients than their possible later identification via the national publication. This issue should be explored further, in terms of whether greater safeguards at the data collection stage should be built in (with possible risk to local ownership and potential delays to viewing the feedback), or whether it is appropriate in this context to dispense with a policy of suppression.

NHS England will explore these issues and potential solutions with information governance experts in order to reach an acceptable solution to the issue of suppression, finding a balance between transparency and a duty of confidentiality.

7.3 Verbatim feedback from the follow-up question

The qualitative review found an appetite amongst patients and the public to be able to view comments from patients in addition to FFT scores for NHS sites. On websites such as Amazon and Trip Advisor, individuals are accustomed to viewing user comments alongside ratings or scores, and participants in the focus groups reported using the comments as a way of evaluating the validity of the corresponding review ratings and coming to an overall judgement about the product or service.

Not all inpatient and A&E sites currently ask a follow-up question, as it was not mandated in the original guidance. Sometimes it is not asked because the methodology used does not allow for it to be directly connected to the response to the FFT question (e.g. where a token system is employed). Some wards publish a selection of comments from the follow-up question on their walls, but these comments are not collected centrally.

One barrier to collecting and publishing comments centrally is the potential burden this would place on the service. This may be considerable, given the scale of FFT, particularly when it is rolled out to all care services. This burden is compounded by the fact that the 80% of FFT responses are presently collected on paper or postcard and would therefore need to be transcribed or scanned. (This presents a strong case for an ultimate shift towards data collection via digital platforms, though care should be taken not to exclude patients less able to complete the FFT in this way.)

A further burden would be the necessity of moderating all comments that are published, anonymising any names referred to, and ensuring that any abusive or offensive language is suppressed. Due to the scale of FFT, this barrier to full publication of comments may be insurmountable, though NHS England could investigate whether publishing a sample of comments would be feasible.

If comments were to be published, then patients would need to be made aware of this and given the opportunity of opting-out. Nevertheless, one risk of seeking to publish all comments is that it could impact on patients' willingness to give honest feedback, both in terms of the overall response rate and of the content of the feedback. This would need to be carefully investigated before implementation, to ensure that FFT's value for service improvement would not be thereby not reduced.

7.4 Publication in departments and wards

As discussed in section 2, the qualitative review found great variation in the extent to which FFT results are publicised in wards. Some wards simply published the FFT score, sometimes alongside other quality measures. Other wards went further, also publishing some of the follow-up comments, or demonstrating how FFT had made a difference through a 'You said, we did' presentation. In some wards, however, FFT results were not shown at all; in these instances ward staff tended to be less knowledgeable about FFT and less engaged in the collection and use of data.

Quantitative FFT data was presented in different ways, with many wards choosing not to use the official FFT score because their staff did not find it salient. The most common alternative was the percentage of patients who said they would recommend the service, which is also how FFT is often reported in local media stories.

At present patients tend not to be informed that their comments may be published, so it is important that safeguards are introduced to ensure that patients understand and have given their consent. As in the case of small numbers, patient anonymity should be preserved, thus any identifiable information should be redacted from published comments. For the same reason, patients' handwriting should not be published, either from the original feedback material or from a scan of it, since this could compromise their anonymity.

8 Conclusions and Recommendations

This section draws together the key findings, outlines the next steps for FFT, and presents the recommendations of this review.

8.1 Ambitions

This review clearly demonstrates that FFT is making a difference to patient experience. 78% say that FFT has increased the emphasis on patient experience in their trust.[33] Its strengths are felt to be the real-time nature of the data, the fact that everyone has the opportunity to feedback and the qualitative data that allows staff to understand both what they are doing well and what they can do to improve the service. There are clearly improvements that can be made to how it is currently being administered and used but its potential to change services is there and should continue to be built upon.

However, when reviewing FFT against its original ambitions, it is clear that it is not meeting these requirements equally. Its value as a tool for local service improvement is currently outstripping its utility for performance management. The question we must ask is whether we can reasonably expect a single way of measuring patient experience to excel in both of these areas. When looking at what is required from patient feedback, we should employ different tools for different purposes and we should not expect one measure to answer all of our requirements.

Different methods of gathering feedback have different strengths. For example, a random probability survey such as the National Inpatient Survey will provide robust data from a representative sample of patients, and is suitable for comparing the performance of different sites. However, it does not offer real-time feedback, as there is a considerable delay between the fieldwork period and when the results are ready to be published. Moreover, it does not provide granular data at ward level. This is not to suggest that the data is poor – it is simply to emphasise that different types of data have clear strengths and related weaknesses. Any methodology has limitations and it is by understanding these limitations that we can draw relevant conclusions about what the data can tell us.

FFT should be considered in the round and continue to be used alongside other ways of measuring patient experience such as the CQC patient survey programme and the GP Patient Survey. The centrally administered surveys should be considered as *summative* measurement tools, designed to take an accurate and comparable reading of performance each year. The FFT, by contrast, should be understood as a *formative* measurement tool, which is designed to promote continual learning and improvement. By playing to the respective strengths of the different programmes we should ensure that we have a patient feedback system that allows us to understand both how to improve services *and* measure progress over time.

The review also found that FFT in its current form is not fully succeeding in informing patient choice about NHS services. This appears in large part due to the present

[33] Ipsos MORI (2014) *The Friends and Family Test: Qualitative research*: pp.1-2. Research conducted on behalf of NHS England.

scoring metric not being salient or understandable to patients, such that they do not know how to assess a hospital based on its present FFT score. The review also found low awareness of FFT among the public. Although the review found that FFT data is not fully comparable, the public may nevertheless wish to view this data in an understandable format such that they can consider it in addition to other indicators of a hospital's quality. As such, in order for its use for public choice to be maximised, FFT data should be presented to the public in a much clearer and understandable manner, albeit one that does not imply direct comparability of FFT data between trusts.

Recommendations for immediate action by NHS England
• There should be clarity of purpose for FFT. The review suggests that there are different understandings of what it is intended to do. NHS England should clearly communicate what FFT can do and how it should be used as described below.
• FFT's clear potential as a tool for service improvement should be emphasised and built upon. Examples of good practice should continue to be collected and disseminated so that trusts can understand from each other what works.
• FFT's role in the patient's right and opportunity to give feedback should be promoted.
• FFT data may help to inform the public, along with other relevant clinical and performance measures, about the quality of NHS services. At present, however, its role as one piece of evidence to help inform patient choice is undermined by the official FFT score, which the public finds difficult to understand. In order to maximise its effectiveness and to raise awareness amongst the public, a transparent and salient presentation of FFT data should be used on NHS Choices, which makes clear that FFT data between different sites is not fully comparable.
• More generally, FFT should not be represented as a statistical measure that is suitable for using to directly compare trusts.
• If financial incentives are to be attached to FFT, then we recommend not basing payments on the relative performance of different sites or trusts.
• FFT should not be seen as a survey. FFT is a new type of patient feedback tool that is offered to all patients and which measures performance continually. So far it has been viewed through the prism of more traditional methods, but this is not helpful. By shoehorning FFT into pre-existing models of data collection and analysis we ask it to perform as something it cannot be, and seen through this lens it will fail.
• Thus far FFT has been categorised as an experimental official statistic. NHS England should work with relevant bodies to establish how FFT should be categorised going forward.

Recommendations for further investigation by NHS England
• Continued work should be done to understand the techniques and methods that can be most effectively applied to both the quantitative and qualitative FFT feedback in order to learn more about what it tells us.
• As FFT is rolled out to other services and the recommendations of this review are put into place a longer term piece of work should look at options to incrementally increase the standardisation of FFT
• FFT shows promise for understanding relative performance within trusts and for monitoring performance over time. NHS England, CQC and local bodies should start to use the data for this purpose
• An insight strategy should be developed by NHS England in order to clarify how FFT fits in with other channels of patient feedback, ensuring that the system comprises a set of complementary tools that respond to different needs and in its entirety provides data that can be used for all the ambitions stated above but does not rely on one to do all them.

8.2 Settings

There is a question mark over the suitability of FFT for A&E. While some A&E staff find it valuable, overall it is not working as well in A&E as in the inpatient setting. There is the practical question of whether FFT can be successfully implemented in such a busy, transitive environment by clinical staff under enormous pressure. There is also a question of whether the FFT question is suitable for the emergency setting, given that there are practical challenges around implementation, as illustrated by the difficulty many A&E departments have experienced in attempting to attain credible response rates. In addition, the public use of FFT data to inform decisions also seems less relevant for A&E.

Recommendations for further investigation by NHS England
• The evidence suggests that we should continue to learn how best to gather patient experience data from A&E, since this poses particular challenges. In particular, many people who are asked at the point of departure in A&E may not have had a chance to reflect on the experience they have received in a clear and lucid fashion.

8.3 Net Promoter Score

This review clearly demonstrates that there are problems with the usage of the Net Promoter Score for presenting FFT results. There is evidence that suggests it provides an incentive to game because only the "extremely likely" category is perceived to be counted positively, and there are stories that patients have, on occasions, been told this by staff.

It is very clear that the current FFT score that is produced from the collected data is widely misunderstood and does not lend itself to clear communication. There are many instances where trusts and local media have used a simple percentage of those who would recommend instead of the NPS because of this issue. If FFT is to become a measure that informs choice and is widely used across the health system then it needs to be simple, intuitive and easy to use. The NPS precludes this and this review recommends adopting another metric.

We do not recommend changing the question or response scale for inpatient or A&E at present, not least because a specific investigation of alternatives has not been conducted. The question and scale are now well-established, and this continuity will help FFT to further bed in. Furthermore, it would be wise to explore the potential utility of the time series data we now have before making any decision that would cause us to lose it. However, negative feedback about both the question and response scale suggests that both might be evaluated against alternatives in the future, if further refinements to FFT are desired.

Consideration should be given to the suitability of the question and response scale in other settings, however, such as the mental health and accessibility for all workstreams.

Recommendations for immediate action by NHS England

• Based on the findings of the qualitative and quantitative reviews, particularly in light of the need to produce a headline metric that frontline staff and the public understand, the NPS FFT score should be replaced by an alternative presentation of the score from one of three viable options:
  ○ A "star system," which converts each FFT response to a star rating between 1 and 5, and which presents the average star rating for the unit (to one decimal place).
  ○ The percentage of people who are likely to recommend the service.
  ○ A positive *and* negative scoring system, which presents both the number who are likely to recommend *and* the number who are unlikely to do so.

• NHS England should work with stakeholders to identify the most appropriate replacement scoring metric and include this in updated guidance.

8.4 Data quality

If it is decided that the key principle of FFT is to provide comparable data to measure relative performance between trusts, then significant changes would need to be made to the data collection. Reliable statistical comparability would require central administration of the test (moving to the model employed by the National Inpatient Survey, the GP Patient Survey and other similar pieces of work). Data adjustment would be required which relies on comprehensive demographic data being available and would necessitate further formalising the sampling process.

There is a key distinction between improving the quality of the data collected in trusts and making it comparable. The findings of this review suggest that there are certainly improvements that can be made to how the data is currently collected, making the data more useful for the service and the public, but that moving to a statistically comparable method would negate many of the key strengths of FFT:

- The data would not be real time;
- We may have to move away from a census approach;
- There would be less local ownership of the findings;
- The cost would increase, particularly if a census approach were to be maintained;
- We would need to limit the collection method to a single mode, most likely paper, losing some of the clear innovation in this area.

Recommendations for immediate action by NHS England
• Do not nationally centralise the collection: the impact on local ownership could undermine the effectiveness of FFT for service improvement, and would not erase question marks concerning comparability (as long as FFT is being administered by the service). This review recommends using FFT alongside the nationally collected datasets, such as the Inpatient survey, rather than asking it to replicate or replace them.
• Take a number of steps to improve the quality of data in order to ensure the feedback is as useful as possible for the service: • This review recommends that tokens should not be used as a data collection method as they have to be collected separately from follow-up comments, meaning scores could not be linked to comments. This reduces the potential utility and meaningfulness of the follow-up comments. • The qualitative data collected after the FFT question is key to the success of FFT. This review recommends mandatory collection of follow-up comments. • There is a balance between collecting demographic data and keeping the simplicity of FFT. The recommendation is that providers should be encouraged to include a limited set of demographic questions where possible. This would allow trusts to determine where there may be selection bias and would allow greater analysis of the data in the longer term. • NHS England should highlight best practice on the timing of the question, explaining the importance, where possible, of doing this consistently within each setting. • NHS England should highlight best practice on the design of data collection materials, with devices that can cause framing, such as images, colours and logos, to be strongly discouraged except where used in materials for audiences that require more accessible designs. • NHS England should highlight best practice on how FFT should be presented to patients to avoid selection bias and framing effects. • Motivations and incentives to game the system should be removed where possible: ○ Stop using the Net Promoter Score to present FFT results ○ Discourage the use of payments attached to relative scores between sites ○ Advise against the construction of league tables
• Develop a support document for the service to understand what they should expect of suppliers, presenting clear standards such as the speed and frequency of providing collated data.
Recommendations for further investigation by NHS England
• There are several different options for the wording of the follow-up question. NHS England should investigate the relative merits of different options, with a view to future standardisation of the question wording.
• NHS England to lead on developing analyses in order to identify and challenge patterns of data that may suggest gaming or other anomalies.
• Explore the role of Healthwatch, CQC and other relevant bodies to hold the system to account on the provision and quality of FFT

8.5 Models of data usage

A clear finding from this review is that much of the success or otherwise of FFT relies on how the data is used within the trust, who has access to it and who sees it as their responsibility to act upon it.

Recommendations for immediate action by NHS England
• Best practice models of data usage learned from the qualitative review and shared, including analysis, dissemination and management techniques, and the engagement of clinicians in FFT.
• NHS England to provide advice on collating data over time in order to get an accurate picture of performance and trends, and on not placing too much weight on small numbers. The inclusion of demographic questions would add value to this process.
• Working models of monitoring and responding to negative feedback to be developed and shared between boards and relevant bodies, including CQC and NHS England.
Recommendations for action by local services
• Trusts to read and assess all of their negative comments as a matter of course. These should be clearly available at board level.
• All data, including follow-up comments, should be made available to all ward staff.
• Front-line staff to be encouraged to review their qualitative feedback. These verbatim comments should have the quantitative question response attached.
• Frequent (weekly) meetings recommended for ward staff to reflect on feedback, celebrate successes, identify problem areas and form solutions.
Recommendations for further investigation by NHS England
• Exploration of ways of giving relevant bodies, including CCGs, CQC and NHS England, access to verbatim comments in order to measure patient experience.

8.6 Data suppression

At present the way that data suppression is currently dealt with on FFT is unsatisfactory. Arguably, it neither guards against confidentiality nor offers transparency.

Recommendations for immediate action by NHS England
• Along with individual response frequencies, overall results should be suppressed for any spatial unit where fewer than five responses have been recorded. This will protect the anonymity of those patients.
Recommendations for further investigation by NHS England
• NHS England to work with relevant information governance bodies to understand the viable options for data suppression and to identify the most appropriate solution, with a view to standardising the policy.

8.7 Patient choice

Modifications to FFT can help make it a more effective tool for informing patient choice. The move away from the NPS presentation will help deliver a metric that is more understandable to the public and which better allows patients to consider FFT data alongside other quality measures to form a judgement about a potential provider of services. Although this improvement will not alter the fact that FFT data is not statistically comparable between trusts, a transparent and understandable presentation should allow patients to decide how to use the data in their consideration of services. Qualitative research found that people use judgement when viewing online reviews and ratings and do not see them as an absolute indicator of quality. FFT should be presented within this context.

There is also a key question of what to do with the qualitative feedback and how this should be disseminated. This is a broader issue facing the system as a whole while FFT rolls out across other NHS services. In particular, the collection of FFT data at GP and Outpatients services will increase the volume of data collection to an industrial scale, such that issue of sustainability may need to be considered, i.e. what can be done within the cost boundary and with the limited resources available to the service. There is a trade-off that needs to be made between ensuring the qualitative feedback is listened to and creating an unsustainable industry around its usage. It is also essential that consent has been given by patients to have their comments made public.

Recommendations for immediate action by NHS England
• Take a number of steps to ensure that FFT data is presented in a prominent, salient and useful manner on NHS Choices: • Move to a new presentation of FFT data, using a new headline metric, and work with NHS Choices to determine best way of presenting this. • NHS Choices to conduct pilot exercises to refine the design of the public-facing presentation of FFT data. • FFT to be published prominently alongside other important clinical and patient experience measures on Choices. • Full FFT response profiles and time series to be included on NHS Choices pages for each institution. • Do not use quintile presentations or any other indicator suggestive of directly comparable statistical data. • Publish the number of responses on NHS Choices and indicate the level of participation this represents in order for users to be able to judge the quality of the data.
• On FFT data collection materials, the FFT follow-up question should always be accompanied by a box that allows patients to opt out of having their comments published in future.

Recommendations for further investigation by NHS England
• Investigate ways of framing data to steer the public away from interpreting data as inferential and comparable, such that they instead treat it purely as descriptive and distinct.
• Investigate, with partners, the publication of anonymised verbatim comments, subject to obtaining patient consent, identifying the risks associated with doing so.

8.8 Communications

At present there is a lack of clarity across the service about the core purposes of FFT. There is also extremely low awareness amongst patients and the public about FFT. In particular, few members of the public know that all patients should have the opportunity to feed back, and fewer still understand how the data is intended to be used. By creating an expectation and an appetite for FFT amongst the public, we can help ensure that implementation of the test remains comprehensive.

Recommendations for immediate action by NHS England
• Communications should go out to the service stating clearly what the purpose of FFT is. Such communications should lay out the principles of FFT that are fundamental to ensuring it makes a difference to patient experience, both in terms of all patients having the opportunity to feed back, and all staff learning from and responding to the feedback.
• Communications should flag the challenge with regard to comparability – that FFT data is first and foremost about the local hospital and that other survey data, such as from the National Inpatient Survey, gives a true picture of inter-trust comparison.
• Communicate to patients that all have the right to provide feedback, and that this feedback is vital for helping the service to improve as well as to help other patients learn about local services when making decisions about care.
Recommendations for further investigation by NHS England
• Explore ways of assessing whether particular patient groups are underrepresented and look at doing specific engagement work with them.

9 Appendices

Please click on each hyperlink to access the appended documents.

9.1 Appendix 1 – The Friends and Family Test: Qualitative research

http://www.england.nhs.uk/wp-content/uploads/2014/07/fft-rev-appx-1.pdf

9.2 Appendix 2 – FFT Review: Quantitative strand

http://www.england.nhs.uk/wp-content/uploads/2014/07/fft-rev-appx-2.pdf

9.3 Appendix 3 – Friends and Family Test Suppressions: A review

http://www.england.nhs.uk/wp-content/uploads/2014/07/fft-rev-appx-3.pdf