Improving Patient Experience in Primary Care: a multi-method programme of research on the measurement and improvement of patient experience

Martin Roland and John Campbell
Co-Chief Investigators

Jenni Burt
Programme Manager

Gary Abel, Ahmed Aboulghate, Faraz Ahmed, Anthea Asprey, Heather Barry, Julia Beckwith, John Benson, Olga Boiko, Pete Bower, Raff Calitri, Mary Carter, Antoinette Davey, Marc N. Elliott, Natasha Elmore, Conor Farrington, Hena Wali Haque, William Henley, Val Lattimer, Nadia Llanwarne, Cathy Lloyd, Georgios Lyratzopoulos, Inocencio Maramba, Luke Mounce, Jenny Newbould, Charlotte Paddison, Richard Parker, Suzanne Richards, Martin Roberts, Claude Setodji, Jonathan Silverman, Fiona Warren, Ed Wilson, Christine Wright
Improve Team

Improving Patient Experience in Primary Care: a multi-method programme of research on the measurement and improvement of patient experience

Martin Roland*[1], John Campbell[2], Jenni Burt[1], Gary Abel[1&2], Ahmed Aboulghate[1], Faraz Ahmed[1], Anthea Asprey[2], Heather Barry[2], Julia Beckwith[1], John Benson[3], Olga Boiko[2], Pete Bower[4], Raff Calitri[2], Mary Carter[2], Antoinette Davey[2], Marc N. Elliott[5], Natasha Elmore[1], Conor Farrington[1], Hena Wali Haque[1], William Henley[2], Val Lattimer[6], Nadia Llanwarne[1], Cathy Lloyd[7], Georgios Lyratzopoulos[1], Inocencio Maramba[2], Luke Mounce[2], Jenny Newbould[1], Charlotte Paddison[1], Richard Parker[3], Suzanne Richards[2], Martin Roberts[2], Claude Setodji[5], Jonathan Silverman[8], Fiona Warren[2], Ed Wilson[1], Christine Wright[2]

[1] Cambridge Centre for Health Services Research, Institute of Public Health, Forvie Site, University of Cambridge School of Clinical Medicine, Cambridge, CB2 0SR, United Kingdom
[2] University of Exeter Medical School, Smeall Building, St Luke's Campus, Exeter, EX1 2LU, United Kingdom
[3] Primary Care Unit, Institute of Public Health, Forvie Site, University of Cambridge School of Clinical Medicine, Cambridge, CB2 0SR, United Kingdom
[4] NIHR School for Primary Care Research, Manchester Academic Health Science Centre, University of Manchester, Williamson Building, Oxford Road, Manchester, M13 9PL, United Kingdom
[5] RAND Corporation, 1776 Main Street, PO Box 2138, Santa Monica, CA 90407-2138, USA
[6] School of Health Sciences, UEA, Norwich Research Park, Norwich, NR4 7TJ, United Kingdom
[7] Faculty of Health & Social Care, The Open University, Walton Hall, Milton Keynes, MK7 6AA, United Kingdom
[8] University of Cambridge Medical School, Cambridge, CB2 0SR, United Kingdom

*Corresponding author

Declared competing interests of authors: Professor Roland and Professor Campbell have acted as advisors to Ipsos MORI, the Department of Health, and subsequently NHS England on the development of the English GP Patient Survey. Pete Bower has received NIHR funding additional to the programme grant. No other authors have any competing interests to declare.

Keywords: primary care, patient experience, quality of care, doctor-patient communication, minority ethnic communities

Abstract

Our research aimed to better understand how different patients use surveys to record experiences of general practice; how primary care staff respond to feedback; and how to engage them in responding to feedback. We used methods including quantitative survey analyses, focus groups, interviews, an exploratory trial and an experimental vignette study.

1. Understanding patient experience data

Patients readily criticised their care when reviewing consultations on video, though were reluctant to be critical when completing questionnaires. When trained raters judged communication within a consultation to be poor, a substantial proportion of patients rated the doctor as ‘good’ or ‘very good’. Absolute scores on questionnaire surveys should be treated with caution: they may present an over-optimistic view of GP care. However, relative rankings to identify GPs who are better or poorer at communicating may be acceptable, as long as statistically reliable figures are obtained. Most patients have a particular GP whom they prefer to see: however, up to 40% of people who have such a preference are unable regularly to see the doctor of their choice. Users of out-of-hours care reported worse experiences where the service was run by a commercial provider compared with not-for profit or NHS providers.

2. Understanding patient experience in minority ethnic groups

We showed that Asian respondents to the GP Patient Survey tend to be registered with practices with generally low scores, explaining about half of the difference in the poorer reported experiences of South Asian compared to White British patients. We found no evidence that South Asians used response scales differently. When viewing the same consultation in an experimental vignette study, South Asian respondents gave higher scores than White British. This suggests that the low scores given by South Asians in patient experience surveys reflect care which is genuinely worse than that experienced by their White British counterparts. We also found that service users of mixed or Asian ethnicity reported lower scores when rating out-of-hours services compared with White respondents.

3. Using patient experience data

We found that measuring how GP-patient communication at practice level masks variation between how good individual doctors are within a practice. In general practices and out-of-hours centres, staff were sceptical about the value of patient surveys and their ability to support service reconfiguration and quality improvement. In both settings, surveys were deemed necessary, but not sufficient. Staff expressed a preference for free-text comments as they provided more tangible, actionable data.

An exploratory trial of real-time feedback (RTF) found that only 2.5% of consulting patients left feedback using touchscreens in the waiting room, though more did when reminded by staff: the representativeness of responding patients remains to be evaluated. Staff were broadly positive about using RTF and practices valued the ability to include their own questions. Staff benefitted from having a facilitated session and protected time to discuss patient feedback.

Limitations

Practice recruitment for our in-hours studies took place in two broad geographical areas which may not be fully representative of practices nationally. Our focus was on patient experience in primary care: secondary care settings may face different challenges in implementing quality improvement initiatives driven by patient feedback.

Future work

Recommendations for future research include consideration of alternative feedback methods to better support patients to identify poor care; investigation into the factors driving poorer experience of communication in South Asian patient groups; further investigation of how best to deliver patient feedback to clinicians to engage them and to foster quality improvement; and further research to support the development and implementation of interventions aiming to improve care where deficiencies in patient experience of care are identified.

Funding

The National Institute for Health Research (NIHR) Programme Grants for Applied Research programme. [607 words]

Contents

List of Tables

List of Figures

List of Boxes

List of abbreviations

ANOVA	Analysis of Variance
APMS	Alternative Provider Medical Services
CAHPS	Consumer Assessment of Healthcare Provider surveys
CAPI	Computer-administered personal interview
CATI	Computer-administered telephone interviews
CFI	Comparative Fit Index
CI	confidence interval
CLRN	Comprehensive Local Research Network
CQC	Care Quality Commission
CRT	Customer Research Technology Limited
DH	Department of Health
DIF	Differential Item Functioning
EQ-5D	European Quality of Life-5 Dimensions
FFT	Friends and Family Test
FTE	Full Time Equivalent
GCRS	Global Consultation Rating Scale
GP	General Practitioner
GPPS	General Practice Patient Survey
HCA	Health Care Assistant
IMD	Index of Multiple Deprivation
IMPROVE	Improving patient experience in primary care
IRT	Item response theory
NHS	National Health Service
NIHR	National Institute for Health Research
NQR	National Quality Requirement
NRES	NHS Research Ethics Service
OOH	out of hours
OR	odds ratio
ONS	Office for National Statistics
PCA	Principal Components Analysis
PCT	Primary Care Trust

PPG	Patient Participation Group
PPI	Patient and Public Involvement
QOF	Quality and Outcomes Framework
RCT	randomised controlled trial
REC	Research Ethics Committee
RMG	Research Management and Governance
RMSEA	Root mean square error approximation
RTF	Real Time Feedback
SD	standard deviation
SHA	Strategic Health Authority
TCC	Test Characteristics Curves
VOP	Value of Patient Feedback scale

Plain English Summary

We aimed to gain a better understanding of how patients in general practice use surveys to record their experiences, to understand how staff respond to feedback, and to find ways of engaging staff more actively in the process. We did this in a number of ways, including carrying out surveys, having discussions with patients and staff, and trying out different ways of gathering patient feedback.

Patients were reluctant to be critical when completing questionnaires after consultations even when they didn't think their care had been very good, and they explained their reasons for this. Considering the results of the national GP Patient Survey, we found that the most common reason for dissatisfaction was patients not being able to see a doctor of their choice – something that has got worse in the past few years.

We carried out several studies to understand why minority ethnic groups give low scores on patient surveys. Part of this is because they tend to be registered in practices with low scores. However, our studies also suggest that low scores from South Asians reflect poor care rather than, for example, different expectations.

We found that, in both primary and out-of-hours care settings, staff don't trust the results of patient surveys and don't usually make big changes in clinical care as a result of these. We looked for ways of engaging staff more actively with patient feedback. We tested 'real time feedback', where patients use a touchscreen in the waiting room, with staff being provided with results fortnightly. Although patients found this useful, few actually used it unless they were reminded by receptionists.

We describe the implications of our research for practice and make recommendations for future research.

[281 words]

Scientific Summary

This programme had seven aims:

1. To understand how general practices respond to low patient survey scores.
2. To estimate the extent to which practice level scores mask differences between individual doctors.
3. To investigate how patients' ratings on questions in the GP Patient Survey relate to actual behaviour by GPs in consultations.
4. To understand patients' responses to questions on communication and seeing a doctor of their choice.
5. To understand the reasons why minority ethnic groups, especially South Asians, give lower scores on patient surveys compared to White British respondents.
6. To carry out an exploratory RCT of an intervention to improve patient experience, using tools developed in earlier parts of the programme.
7. To investigate how the results of the GP Patient Survey can be used to improve out-of-hours care.

In this report, we group our results under three headings:

- Understanding patient experience data (aims 3 and 4)
- Understanding patient experience in minority ethnic groups (aim 5)
- Using data on patient experience for quality improvement (aims 1, 2, 6 and 7)

We conducted empirical studies in GP practices (varying in location, deprivation and performance on patient experience measures) and out-of-hours providers nationally. A total of 47 GP practices and 11 out-of-hours providers participated in the programme of work, although some were involved in more than one study. We additionally completed multiple analyses of GP Patient Survey data and, for an experimental vignette study, collected data from the general public.

1. Understanding patient experience data

Patient surveys are widely used in many countries, yet comparatively little is known about what experiences actually lead patients or service users to respond in particular ways when completing them. We approached this issue in two studies, in which we (1) used video elicitation interview methods to ask 52 patients directly about how they chose certain questionnaire responses while showing them a video of their consultation and (2) used statistical analyses to compare assessments of videoed consultations by 56 patients with those of expert raters using standardised assessment instruments.

The first study (chapter 2) showed that while patients readily criticised their care when reviewing consultations on video, they described how they had been reluctant to be critical when completing a questionnaire. Reasons for this included the need to maintain a relationship with the GP, gratitude for NHS care they had received in the past, and power asymmetries. We concluded that patients find questionnaires to be limited tools for feeding back concerns about consultations.

The second study (chapter 3) reinforced our conclusion from chapter 2. When trained raters judged communication in a consultation to be good, patients generally did the same. However, when trained raters judged communication in a consultation to be poor, patients' assessments varied from poor to very good.

The tendency for patients to choose positive responses suggests absolute survey scores should be treated with caution: they may present an over-optimistic view of the GP's care. Surveys can be used to look at relative scores: scores from a GP which are lower than comparable practices are likely to indicate a problem.

We spoke to GPs about their survey results (chapters 7 and 8) through both focus groups and face-to-face interviews: they reported how, whilst positive about the concept of patient feedback they struggled to engage with and make changes under the current approaches to measurement.

A second aspect of care which we identified as being of importance as part of our programme of work relates to patients' ability to see a doctor of their choice. Our analyses of GP Patient Survey data showed that at all age groups most patients have a particular GP whom they prefer to see. However, up to 40% of people who have such a preference are unable regularly to see the doctor of their choice. This is a significant quality issue for the NHS.

In out-of-hours care, we found that patients reported worse experiences where the service was run by a commercial provider compared with not-for-profit or NHS providers. We discuss possible reasons for this.

2. Understanding patient experience in minority ethnic groups

Minority ethnic groups provide consistently low scores in English surveys. In this theme, our analyses of GP Patient Survey data focused on South Asian respondents and on questionnaires completed in English. Although the GP Patient Survey is available in 15 languages, fewer than 0.2% of surveys are completed in languages other than English. First we showed that Asian respondents to the GP Patient Survey tend to be registered in practices with generally low scores, explaining about half of the difference between South Asian and White British patients in their experience of care (chapter 5). Then, using item-response theory, we found no evidence that South Asians used the scales in a different way to White British respondents (chapter 5).

We then conducted an experimental vignette study, for which we filmed 16 simulated consultations based on transcripts of real consultations using various combinations of White and Asian doctors and patients, half scripted to be ‘good’ and half ‘poor’ for communication. We showed three randomly sampled videos to each of 1120 people (half White British, half Pakistani, equally split between those under and over 55) and asked them to score the consultation using the communication items from the GP Patient Survey. If the low scores of Pakistani patients in real life settings were due to higher expectations of care, then we would expect them to give lower scores in the experimental vignette situation. In fact, the reverse was observed. When viewing the same consultation, Pakistani respondents gave scores which were much higher when adjusted for socio-demographic characteristics, compared to White British respondents. This suggests that the low scores given by Pakistani patients in surveys such as the GP Patient Survey reflect care which is genuinely worse, and possibly much worse, than that experienced by their White British counterparts.

It is sometimes suggested that survey scores should be adjusted for the ethnicity of the respondents in order to compare practices with high or low proportions of minority ethnic patients. Our results suggest that this should not be done: rather, low scores from South Asian patients should be taken at ‘face value’ and investigated as possible indicators of poor care.

3. Using data on patient experience for quality improvement

In chapter 9, we summarise a patient experience survey we conducted to explore assessment of care conducted at practice level compared to individual GP level. In chapters 7 and 8 we describe focus group and interview studies in which we sought the views of GPs and practice staff on survey results. In chapter 11 we describe a qualitative interview study with out-of-hours staff responsible for collecting and acting upon patient feedback which explored the same issues, alongside other work on measuring patient experience of out-of-hours care.

By conducting a patient experience survey at individual doctor level, we demonstrated that practice-level ratings of GPs’ communication skills can mask considerable variation between GPs within a practice. This is particularly the case in poorer performing practices, where patients may experience wide variation in communication skills between individual doctors.

Across both settings, staff neither believed nor trusted patient surveys. Concerns were expressed about their validity and reliability and of the likely representativeness of respondents. Staff expressed a preference for free-text comments as they provided more tangible, actionable data. It was easiest for practices to engage with office functions such as appointment systems and telephone answering. Addressing an individual doctor’s performance (e.g. communication skills) was much more difficult.

In interviews, doctors expressed markedly ambivalent views about surveys. However, despite their concerns about surveys, they expressed broadly positive views about the importance of patient feedback in monitoring and improving services.

These results led us to consider how patient feedback might be obtained and delivered in a way that would engage doctors to use patient surveys for quality improvement. We conducted a preliminary evaluation of a real-time feedback (RTF) based intervention in general practice, using touch screens that patients could use to leave feedback following a consultation. Since RTF has not been widely used, an exploratory RCT and qualitative study were conducted to answer questions about feasibility, to estimate likely response rates, to get patient and staff views on RTF, and to estimate the costs to a practice of introducing RTF. We also included facilitated feedback in one arm of the exploratory trial. In our exploratory trial, only 2.5% of consulting patients left any RFT without prompting; however, if encouraged to use RTF by staff, as many as 60% of patients did so. Of patients who used RTF, 86% found it easy to use and were positive about it as a feedback method. Lack of awareness of the screens and lack of time were the commonest reasons for not giving feedback.

Practice staff were broadly positive about using RTF and practices valued the ability to include their own questions in the survey. Practices which had open communication between staff members tended to be more positive about using patient feedback. Practice staff identified clear benefits from having a facilitated session for discussion of patient feedback and having protected time to discuss the results.

Our programme of work was supported by two study advisory groups comprised of lay members and health care professionals: one, based in Cambridge, which provided support across all streams of work except the out-of-hours research, and one, based in Exeter, convened specifically to provide input to the out-of-hours workstreams.

Implications for practice

The work that we have carried out over the past five years has clear implications for practice.

1. The importance of patient experience

Our research supports the continuing emphasis on obtaining patient experience feedback as an important means of informing NHS care. Whilst continuing effort should be invested in refining the most effective and meaningful mechanism to capture patient feedback, the key challenge remains to provide primary care staff with the support and means to enable them to act upon patient feedback.

2. The need for action on the quality of care for minority ethnic groups

There has been much speculation as to whether the lower survey scores reported by minority ethnic groups are “real”, reflecting poorer quality of care, or an artefact of the questionnaires used or higher expectations of care. Our series of studies strongly suggests it is the former, with patients from South Asian backgrounds experiencing considerably poorer communication with GPs than their White British counterparts. Effort should be invested to ensure lower scores on patient experience surveys from such groups are investigated as markers of poorer quality of care.

3. Patients give over-positive responses when rating their care

Our results highlight the difficulty that patients have in feeding back negative experiences in questionnaire surveys. However, patients' reluctance to criticise a doctor or provider with whom they have to maintain an ongoing relationship will not be addressed simply by changing the survey method. Providers and managers need to understand that absolute scores paint an optimistic picture of patients' true views.

4. Surveys are not sufficient to fully capture patient feedback

Across primary and out-of-hours care settings, staff view patient surveys as necessary, but not sufficient. Alternative methods for gaining more qualitative feedback were commonly used to supplement survey scores, with free text viewed as providing more actionable data than responses to standard survey questions.

5. The need for individual level feedback for doctors

Reporting patient experience at practice level masks substantial variation in performance within practices for aspects of care related to individual doctors (e.g. doctor patient communication). However, if a practice has overall high scores for doctor-patient communication, it is unlikely that such a practice contains a low scoring doctor. Robust mechanisms are needed to help lower scoring practices identify and support doctors whose individual patient feedback identifies areas for potential improvement.

6. Patient surveys need to become more meaningful to staff

Practices found it easier to engage with items on surveys that related to practice management (e.g. availability of appointments, ability to get through on the phone) than to issues around communication between patients and clinical staff. Effort should be invested to focus the attention of staff on the whole range of feedback provided by service users, and on making available suitable support and learning opportunities to act on such feedback.

Immediacy of feedback, regularity of feedback, and having some control over the questions asked were all aspects of our experiment with RTF that were valued by practices and had the potential to make feedback more useful. However, a number of important questions remain before RTF could be recommended as a replacement for postal questionnaires.

7. The value of surveys in monitoring national trends

Despite their limited value in stimulating quality improvement, surveys are important for monitoring national trends. For example, the GP Patient Survey is the only source of data which demonstrates that, year on year for the past five years, patients report that they have had increasing difficulty in seeing a doctor of their choice. For out-of-hours services the GP Patient Survey is the only way of monitoring trends. However, where national surveys are used to monitor trends in care it is important that the questions (such as questions on access in the GP Patient Survey) do not keep changing. Much smaller sample sizes are required to monitor national trends: tens of thousands of participants rather than millions.

8. Development of surveys in out-of-hours care

Out-of-hours services are required to audit patient experience but are provided with no information on how to do this, leading to diverse and non-comparable data. Our work shows that, subject to minor amendments, the GP Patient Survey *is* suitable for the national monitoring of OOH care. However, it is not sufficiently detailed to support quality improvement, and is unlikely at present to replace in-house approaches, leading to duplication of effort.

Concluding remarks

Large scale postal surveys are likely to remain the dominant approach for gathering patient feedback for the time being, although a range of other methods are being developed. These include real time feedback, focus groups, online feedback, analyses of complaints, practice participation groups, and social media. In the final section of our report we outline recommendations for research, and identify the criteria that any new methods will need to meet in order to become useful quality improvement tools.

[2369 words]

Chapter 1. Introduction to the IMPROVE programme

Context

Improving the health status of individuals and populations is a central ambition of western healthcare systems, and the US Institute of Medicine has suggested that high quality healthcare delivery should be safe, effective, patient-centered, timely, efficient and equitable.[1] Berwick et al. [2] have recently noted the importance of patient experience of care as one of the suggested ‘triple aims’ of an advanced healthcare system. A recent US report highlighted the important contribution that listening to, and acting on, patient feedback can potentially make to healthcare improvement efforts. [3]

New developments within the English NHS highlight the embedding of public performance assessment within the regulation of the health care system, including NHS England’s consultation on the production of GP league tables, [4] and the Care Quality Commission’s parallel development of a rating system for primary care. [5] A transparent health care system is regarded by policy makers as essential to enabling patients to make informed choices about the care they receive, [6] and patient feedback on healthcare services is now commonly gathered in the US, Canada, Europe, Australia, and China.

Efforts to improve quality of care in the NHS over the last 15 years have focused on providing prompt access to care (for example, the time taken to see a GP or hospital waiting times), and on providing evidence-based clinical care (for example through the development of National Service Frameworks and the Quality and Outcomes Framework). A direct link between patient feedback and quality improvement efforts was previously operationalised by including results arising from patient surveys as a component of the UK Quality and Outcomes Framework (QOF). [7] This performance management system provides financial incentives for GPs within the NHS to achieve agreed quality indicators covering areas including chronic disease management, practice organisation and additional services offered. With the introduction of QOF it was possible, for the first time, to rank all practices according to their patient feedback, and results of surveys, aggregated at practice level, formed the basis of a pay for performance scheme between 2009 and 2011, when the UK government withdrew the pay for performance arrangements for patient experience.

Some of these policies have been highly effective – for example, associated with a wide range of quality improvement initiatives over a decade, there have been greater improvements in the UK for the clinical care of conditions like heart disease and diabetes than in any other major developed country. [8] Although relatively neglected in the early years of the millennium, patient experience of healthcare is now a high policy priority, and in 2008 the Next Stage Review [9] suggested that:

"…..quality of care includes quality of caring, this means how personal care is – the compassion, dignity and respect with which patients are treated. It can only be improved by analysing and understanding patient's satisfaction with their own experiences."

The review however noted that "[up until 2008] progress has been patchy, particularly on patient experience" and announced the development of quality accounts for all NHS organisations in which "….we will require healthcare providers to publish data … looking at safety, patient experience, and outcome."

Since 2008, therefore, there has been a major policy initiative to improve patient experience in the NHS. Most recently, the focus on patient experience has been enshrined in the NHS Outcomes Framework which, in Domain 4, focuses on ensuring that 'patients have a positive experience of care'. [10] In primary care, these policy initiatives and statements have been implemented primarily through the development and conduct of the GP Patient Survey (www.gp-patient.co.uk), first sent to 5.6 million patients in January 2009. The large sample size was intended to provide sufficient responses to characterise patient experience of primary care in all 8,300 GP practices in England. Detailed responses for individual practices were published on the NHS Choices website (www.nhs.uk) and made available online and included information on access to GP services and on interpersonal aspects of care, out-of-hours care, and care planning. The questionnaire specifically included validated questions about inter-personal aspects of care, based on questionnaires which the authors of this present report previously designed and on which we have previously reported. [11] This large-scale survey is, of course, an expensive undertaking, and the utility and impact of the GP Patient Survey needs to be commensurable with this investment.

In seeking to achieve improvement in the quality of NHS services, gathering data is important both to inform the process of service development and innovation, and to assess the impact of such changes in practice. It has been suggested that data to support such improvement initiatives needs to be of sufficient quality to assess whether an innovation can be made to work, rather than the more rigorous level of research data needed to assess whether an innovation works. [12]

Communication in the consultation has always been an important part of primary care and is closely linked to continuity of care. At the outset of this research, there had been many anecdotal accounts that GPs were more focused on meeting clinical targets identified on their computer screens than on the needs of the patient sitting in front of them. It seemed therefore an appropriate time to balance the focus on improving clinical care with a renewed focus on inter-personal care and on communication in the consultation. The ability of patients to choose their own doctor is also important. Our research prior to commencing this programme showed that continuity of care had deteriorated since the introduction of the new GP contract in 2004, [13] and previous research had also highlighted that patients were less likely to report overall positive experiences if they were not able to choose a doctor whom they know. [14, 15]

Experience and satisfaction

Previous research has identified considerable confusion and overlap relating to the concepts of patient experience and satisfaction. The two concepts are closely linked, although at a simple level, reports of experience relate to recounting or commenting on what actually happened during the course of a clinical encounter, whilst reports of satisfaction focus on the patient or carer’s subjective evaluation of the encounter (that is, asking for “ratings” of care rather than simple “reports” of care). Individual items in a survey may thus examine patient’s reports of their experience of care, whilst other items may explore patient’s evaluation of that care, the linkage between report and evaluation/rating item pairs offering potential for the development of cut points in scales of performance. [16] In practice, however, the terms are often used interchangeably and survey items designed as report items often have an evaluative component, for example the question “Were you involved as much as you wanted to be in decisions about your care and treatment” from the NHS Inpatient Survey contains elements of both. Within the GP Patient Survey, the instrument behind much of this programme of work, items often relate to ratings of care. For example, the communication questions ask patients to consider “how good” the doctor was at providing various elements within a consultation, including giving enough time, involving in decisions about care and treating with care and concern.

Patient satisfaction may be seen as a multidimensional construct, focusing on the subjective experiences of patients, and related to their expectations of care, and the perceived technical quality of the care provided. [17] Russell [18] has recently summarised some of the problems associated with surveys of patient satisfaction with care, including problems with the validity and reliability of satisfaction survey instruments, the lack of a universal definition of the term satisfaction, the disinclination for patients to be critical of care received because of not wanting to jeopardise their treatment, satisfaction being determined by factors other than the actual health care received, and the frequently non-specific nature of resulting findings arising from such surveys. In contrast, reports and surveys of patient experience may offer potential to discriminate more effectively between practices than do reports of patient satisfaction [19] thus potentially offering greater external accountability of health care providers, enhanced patient choice, and greater potential to improve the quality of care and to measure the performance of the health care system as a whole. [20]

Patient experience matters

Patient experience is an important end point for NHS care in its own right. Patients consistently report that personal care is central to effective care, and in that context, the development and refinement of GP’s interpersonal skills are a key priority. [20, 21] It is noteworthy that many complaints regarding care centre not on technical and ‘clinical’ aspects of care, but on issues relating to inter-personal aspects of care and communication. [22, 23] Good communication with patients is not just an end in its own right; it brings three important additional benefits.

Firstly, our research [24] has confirmed earlier work which showed that patients balance a range of beliefs and concerns when making decisions about taking medicines. Adherence is related both to the quality and duration of the consultation, and to the doctor’s ability to elicit and respect the patient’s concerns. [25-27] Better communication may lead to improved patient outcomes [28] through, for example, improved blood pressure control in hypertensive patients.[29]

Second, there is a close relationship between poor communication and serious medical error.[30] This is partly because not listening to the patient's perspective may lead doctors to miss important clinical information, and partly because patients react more negatively when things go wrong if communication has been poor during the clinical episode in question. A significant proportion of cases referred to medical defence societies have at their heart poor communication in the consultation, [31] and improving communication with patients and engaging them more closely in their care is seen as key to improving patient safety. [32, 33]

Third, the increasing emphasis in the NHS on self-care and prevention demands good information and shared decision making in the consultation. Our research shows that GPs and practice nurses are currently poorly prepared for roles in which they encourage patients to take greater responsibility for their own care [34] or their lifestyle choices.

Although intuitively of importance, enhanced patient experience of care also matters on account of an important range of other associations reported in the research literature, including improved safety-related outcomes, [35] self-reported health and wellbeing, [36] enhanced recovery, [28] increased uptake of preventive health interventions, [37, 38] and reduced utilization of health care services including hospitalization and emergency department visits.[39]

Capturing patient experience of care

Whilst several approaches have been adopted to obtaining information on patient experience of care – for example through the use of focus groups, patient participation groups, in depth patient interviews, feedback booths placed in healthcare settings, direct observation of patient experience, [40] and the use of compliment and complaint cards to capture qualitative feedback – the only practical approach to capturing large-scale feedback with the intent of providing actionable information remains through the use of surveys of patients. In primary care in England, this culture of feedback has been embedded into routine practice in several ways. Central amongst these is the use of structured patient feedback obtained through surveys of patients' experience of care, both at national and practice levels. [41]

Qualitative approaches may be judged to offer greater depth of feedback than quantitative approaches, [42] but such approaches are intensive in respect of data collection, although

Locock et al. [43] have drawn on secondary analysis of a large national qualitative data archive to inform service improvements.

Newer forms of capturing feedback, such as the use of tablets and kiosks to capture real time feedback is an area of great current interest, but ones which, as yet, lack a strong evidence base from primary care. During the course of this research, a report from a preliminary observational study [44] suggested that real-time feedback offers potential in primary care settings, and similar findings [45] have emerged from reports provided by patients with cancer attending oncology out-patients. Whilst there may be potential for widespread use of real-time data capture of patient experience in primary care, the acceptability and feasibility of the approach in routine primary care is not known, and nor is the nature of the feedback provided. Such an investigation needs detailed feasibility and pilot work using an experimental design of real-time feedback of patient experience of primary care.

Large scale surveys of NHS patients and staff have been in use since the mid-1990s, building on the experience of smaller-scale surveys conducted at local level, or on the experience of surveys conducted for research purposes. Large scale surveys of patient experiences of primary care were first introduced in 1998 [46] with the express purpose of addressing issues relating to the quality of care and reducing inequalities in care by taking account of patient's views in informing local service developments. Surveys of patients have been used extensively since the introduction of the UK Quality and Outcomes Framework in 2004, when two questionnaires (GPAQ [11]and CFEP [47]) were 'approved' for use by the NHS, and adopted as the basis of linking the pay of general practitioners to their participation in the patient survey programme. [48]

Such surveys may be administered in a variety of ways. In healthcare contexts, paper-based surveys are most commonly used, although digital e-platforms are now commonly and widely used as a means of capturing information, most frequently using online processes. Computer-administered personal surveys (CAPI) [49] and computer administered telephone interviews (CATI) may also be used, most commonly in research settings.

The NHS has established a major programme of surveys [50] developed for a wide range of settings. Several of these surveys focus on patient experience of care, emulating the suite of

Consumer Assessment of Healthcare Provider surveys (CAHPS) introduced in the US in 1995. [51]

The content of primary care surveys of patient experience

Historically, the content of UK primary care surveys has evolved from the 1998 survey, [52] which covered a wide range of issues including primary care access and waiting times, GP-patient communication, patients' views of GPs and practice nurses in terms of knowledge, courtesy, and other personal aspects of care, and the quality and range of services provided such as out-of-hours care and hospital referrals. The GP Patient Survey in 2008 developed and presented an expanded suite of items from the surveys of 2006 and 2007, which were focused almost exclusively on the accessibility of GP services; the 2008 survey focused on domains of care identified as being of importance to patients, [53, 54] including the accessibility of care, technical care, inter- personal care, patient-centredness, continuity, outcomes, and hotel aspects of care. More recently, the English National Health Service (NHS) has outlined eight domains believed to be of critical importance in respect of patient experience. Overlapping with earlier thinking, these include respect for patient-centred values, information, communication and education, emotional support, physical comfort, continuity, and access to care, [55] all being reflected, at least to some extent, in the ongoing GP Patient Survey programme.

Most recently, the 'Friends and Family Test' has been introduced widely across the NHS, acting as a single-question proxy for patient experience based on the willingness of respondents to recommend their healthcare provider to close acquaintances. The widespread use of the test has been accompanied by specific guidance on its implementation in practice, [56, 57] and research reports have recently started to emerge following the use of the test in hospital settings, in which concerns have been raised about the reliability of the test. [58, 59] The use of the test was rolled out to GP settings in December 2014.

Out-of-hours services

Beyond the domains mentioned above, additional areas of enquiry incorporated in the 2008 version of the GP Patient Survey included out-of-hours care and care planning. Variation in patients' experience of out-of-hours care has been identified as an area of concern since 2000, with numerous influential reports considering the structures suitable for delivering out-of-hours care, as well as highlighting the variable experience of patients across the UK in respect of service delivery. In 2000, Dr David Carson reported on the structural aspects of out-of-hours care pertaining at the time, and recommended an expanded role for NHS Direct as a facilitator of access to these GP-led services, proposing that patients should use a single telephone access point to enter the system. [60] Much less emphasis was placed on patients' experience of out-of-hours care, although recommendations were made regarding the need to monitor patients' experience of the developing service. The transfer of responsibility for out-of-hours care from GPs to primary care trusts was foreshadowed in a report of the House of Commons Health Committee [61] which once again focused on structural and organisational issues relating to care provision. It was not until 2006, [62] following the publication of national quality requirements in respect of out-of-hours care in October 2004, [63] that patient experience of such services began to attract serious attention, with a recognition that, by 2006, although patient experience of out-of-hours services was generally 'good', 1 in 5 patients were dissatisfied with the service at that time, and 40% of respondents in an independent survey of out-of-hours service users reported that the overall quality of the service was less than 'good'. [64] The incorporation of six items in the 2008 GP Patient Survey with the intent of capturing information on aspects of out-of-hours GP services thus represented an extension of earlier versions of the questionnaire, recognising the growing importance of patient experience of care, and offered the potential to examine the experience of patients from various sub-groups of the population, and the potential to compare out-of-hours service providers in respect of their patients' experience of care.

Measuring patient experience of care

The potential utility of questionnaires capturing patient feedback is, like other questionnaire-based feedback, dependent on the psychometric performance of the questionnaire in practice. Issues centering on the validity of the resulting data – whether the questionnaire items are measuring what is intended to be measured rather than some extraneous domain – underpins the reliability of inferences and conclusions which might be drawn following data collection. Validity and reliability themselves each comprise several elements, and demonstrating validity of an assessment is generally regarded as having priority over demonstrating reliability.

Our previous research identified concerns expressed by doctors regarding the use of patient survey data for the purposes of providing individual feedback regarding a doctor's performance. [65] Some of those concerns focused on the reliability and validity of the resulting data and on the conclusions being drawn regarding a doctor's professional practice.

Validity of items within a questionnaire may be assessed in a number of ways – for example in exploring the pattern of item response using quantitative approaches such as factor analysis to investigate the latent variables identifiable within the theoretically related item-responses. Qualitative approaches may also be used in the questionnaire design phase – for example, cognitive interviews were undertaken with patients from a range of socio-demographic backgrounds in the early stages of developing the GP Patient Survey. [66] Such interviews are designed to assess the interpretability and accessibility of the putative questionnaire items. Qualitative studies using cognitive interviewing or similar approaches may, however, be undertaken following respondent's completion of questionnaires, seeking to explore the basis on which respondents are providing their evaluation. Such studies are unusual, but potentially offer great value in exploring respondent's insights, and exploring whether items presented are interpreted as originally intended.

Patient's varying experiences of care

Our earlier research, and the research of others, has previously identified substantial variation between practices in patients' reports of their experience of, and satisfaction with, care received. Recent studies have also identified a range of patient experience being reported amongst doctors providing care in similar clinical specialties and settings. This acted as the basis for the inclusion of patient feedback as an element required by UK regulatory authorities for the routine appraisal and revalidation of doctors. Despite these observations, few studies have examined the relationship between feedback on patient experience aggregated at practice level and the performance of individual doctors within practices, with one observational study [67] identifying a substantial range of performance amongst doctors from a sample of eight Scottish general practices; the authors noted a number of possible contributing factors which might have accounted for differences observed at doctor level, including the experience of the doctors themselves, as well as the doctor's mental health and professional disillusionment.

Systematic differences in patients' reports of their experience of care have also been reported to be related to the characteristics of patients themselves. Older patients, patients from white ethnic backgrounds, the better-educated, the less deprived, and those reporting better health status have generally reported more favourable experiences of care when compared with younger patients, those from minority ethnic groups, the less well educated and more deprived, and those with poorer health status. Similar differences have been reported across many healthcare systems, and have given rise to calls to take account of the characteristics of participating patients when considering the results of patient feedback on care. To date, however, such calls have generally not been heeded in practice, since the relative contribution of practice, doctor, and patient to overall variation in feedback remains to be defined. Uncertainty regarding the need for, and effect of, such 'casemix adjustment' remains a concern for doctors in their consideration of patient feedback.

Specifically in respect of variation in experience amongst patients from different ethnic backgrounds, previous analyses have identified variations in patient experience in relation to ethnic group, age, and gender, and have found an interaction between ethnicity and age for cancer referrals. [68, 69] However, no studies to date have yet investigated such an interaction in respect of patients' experience of communicating with their GP – for example, investigating differences between older and younger patients, by gender, amongst patients representing a range of minority ethnic groups.

In addition, although communication between doctors and patients is a core component of patient experience, [70] and minority ethnic groups have reported lower patient experience scores for communication compared to the majority population, [68, 71, 72] such differences are not consistent for all minority ethnic groups. Previous analysis of patient experience data conducted by the authors highlighted that South Asian patients reported particularly negative experiences, including for waiting times for GP appointments, time spent waiting in surgeries for consultations to start, and continuity of care. [68] However, such analyses have not been repeated using GP Patient Survey data.

A number of potential explanations have been suggested for the lower ratings provided by South Asian and other minority ethnic groups in respect of their experience of care. Broadly, these relate to whether South Asian patients (a) receive lower quality care, or (b) receive the same care, but rate this more negatively. [68] For example, differences in the use of questionnaire response scales might lead to South Asian groups being less likely to endorse the most positive options when asked to evaluate a doctor's communication skills. [73] Alternatively, there could be systematic variations in evaluations of consultations because South Asian respondents vary in their expectations of, or preferences for, care. However, recent evidence from the US points to lower quality of care as the main driver of variations. [74] Gaining understanding of why minority ethnic groups give relatively poor evaluations of their care is key to forming an effective response, as determining appropriate action is difficult until it is ascertained whether differences in evaluations relate to true differences in care or to variations in expectations, scale use, and preferences. Exploring these observed differences between patients from various ethnic backgrounds is challenging using only observational, real-world, data – more robust approaches are required, drawing on experimental designs in which some key elements of the consultation–interaction can be accounted for in the analysis – for example through the use of standardised consultations and video-vignettes. [74, 75]

Using patient survey data to improve care

Although there is a belief, articulated in the Darzi Review, that patient surveys can be used to improve care, a systematic review from 2008 suggested that there is considerable uncertainty about how and whether this can actually be achieved. [9, 76] Several causal pathways for achieving improvements in provider performance through the release of publicly-reported performance data have been proposed. [77-79] Some invoke market-like selection, claiming that patients will modify their choice of provider using publicly available data, such as that provided by patient experience websites. [77, 80-82] Evidence to support this pathway is, however, weak. [78] A more likely mechanism driving performance improvement in response to the publication of performance data is health professionals' concern for reputation, in which peer comparison motivates individuals and organisations to improve their care. [78, 79]

Furthermore, at the outset of this research, Primary Care Trusts were poorly prepared to support and work with GP practices to improve patient experience. In addition, the Darzi Review had noted that progress in improving patient experience in the NHS had been slow, [9] and our research had identified that some aspects of care, especially out-of-hours care [83, 84] and continuity of care, [13] may actually have worsened in recent years. In addition, as observed above, it had been noted that patients from minority ethnic communities consistently reported lower evaluations of the quality of primary care. [68] Although these problems had been clearly identified in published research, the research had provided less clarity about the meaning and interpretation of these findings and the best way to intervene to improve patient experience.

Irrespective of its potential to stimulate change, the publication of performance data is central to the openness and transparency that are seen as essential to a safe, equitable, patient-centred health care system. [85] Thus, regardless of any effect on quality improvement, such initiatives are likely to be here to stay. [78] In refining the information made public, it is important that performance data are accurate and relevant to all potential users. The US-based Robert Wood Johnson Foundation has noted that while patients "prefer to see comparative information for individual providers rather than practices or groups", this is often not done in practice. [86] Currently, there is some move towards publication of performance data from an organisational level to that of individual doctors. In the UK, for example, patients referred to the cardiology service at the South Manchester Hospitals Trust may go online to view both mortality and patient experience data for each cardiologist or cardiac surgeon. [87] However, within English primary care, the practice-level aggregation of data from the GP Patient Survey used to derive practice performance indicators potentially masks variation in performance amongst individual GPs, thereby inappropriately advantaging or disadvantaging particular doctors. Current indicators may consequently fail to provide users, providers or commissioners with an accurate assessment of performance within a practice.

Although intuitively simple, patient satisfaction is a complex concept [88] and patient responses to questionnaire scores must be interpreted carefully. For example, practices need to understand if low ratings of communication reflect particular consultation behaviours, or whether they are in fact due to broader issues such as practice culture, or the structure and availability of consultations and appointments.

Once the causes of low ratings have been better understood, interventions to improve care can then be designed. However, the current literature on the effects of feedback of patient assessments is insufficient in scope, quality and consistency to design effective interventions.[76] There are many reasons why simple feedback on patients' experience of care is likely to have limited effects. Our research is designed to address these gaps in knowledge, to enable managers, patients and professionals to have confidence in the meaning of patient assessments, and to provide effective interventions to improve care when problems are identified.

Summary

In summary, therefore, capturing patients' experience of primary care is a current ambition of major importance in UK government health policy. Patient surveys, incorporating opportunities for people to comment on various aspects of their care, are a convenient means of capturing relevant information at scale. It is not clear, however, how healthcare staff operating in practices respond to the resulting information. Previous experience suggests that staff may rationalize scores on the basis of concerns regarding the scientific properties of the survey, or uncertainty regarding the implications arising from providing care in their particular circumstances – for example, taking account of the socio-demographic mix of respondents. On a similar vein, it remains unclear the extent to which overall practice performance, based on aggregated patient feedback, might relate to the performance of individual doctors within the practice. It is also unclear whether patients provide reliable evaluations of care – and the extent to which such evaluations might vary according to the socio-demographic characteristics of respondents. New modes of capturing patients' experiences of their care have become available in recent years, but to date, it is not clear whether novel, technology-based approaches can be successfully implemented in routine primary care settings, nor the extent to which any resulting data might reflect the results of the wider population.

In recent years, care provided by out-of-hours GP services has been a particular area of interest for the NHS, and has been the subject of national audit and standard-setting. However, it is not clear whether patients' reports of their experience of out-of-hours care are valid and reliable. Neither is it clear the extent to which factors relating to the structure and organization of such care might be associated with systematic differences in patients' reports of their care. Furthermore, as for in-hours care, it is not clear how staff providing out-of-hours care might respond to patient feedback, and how service managers might utilize such information in the planning and design of services aimed at being responsive to the needs of NHS patients.

Aims of the programme

This programme had seven aims:

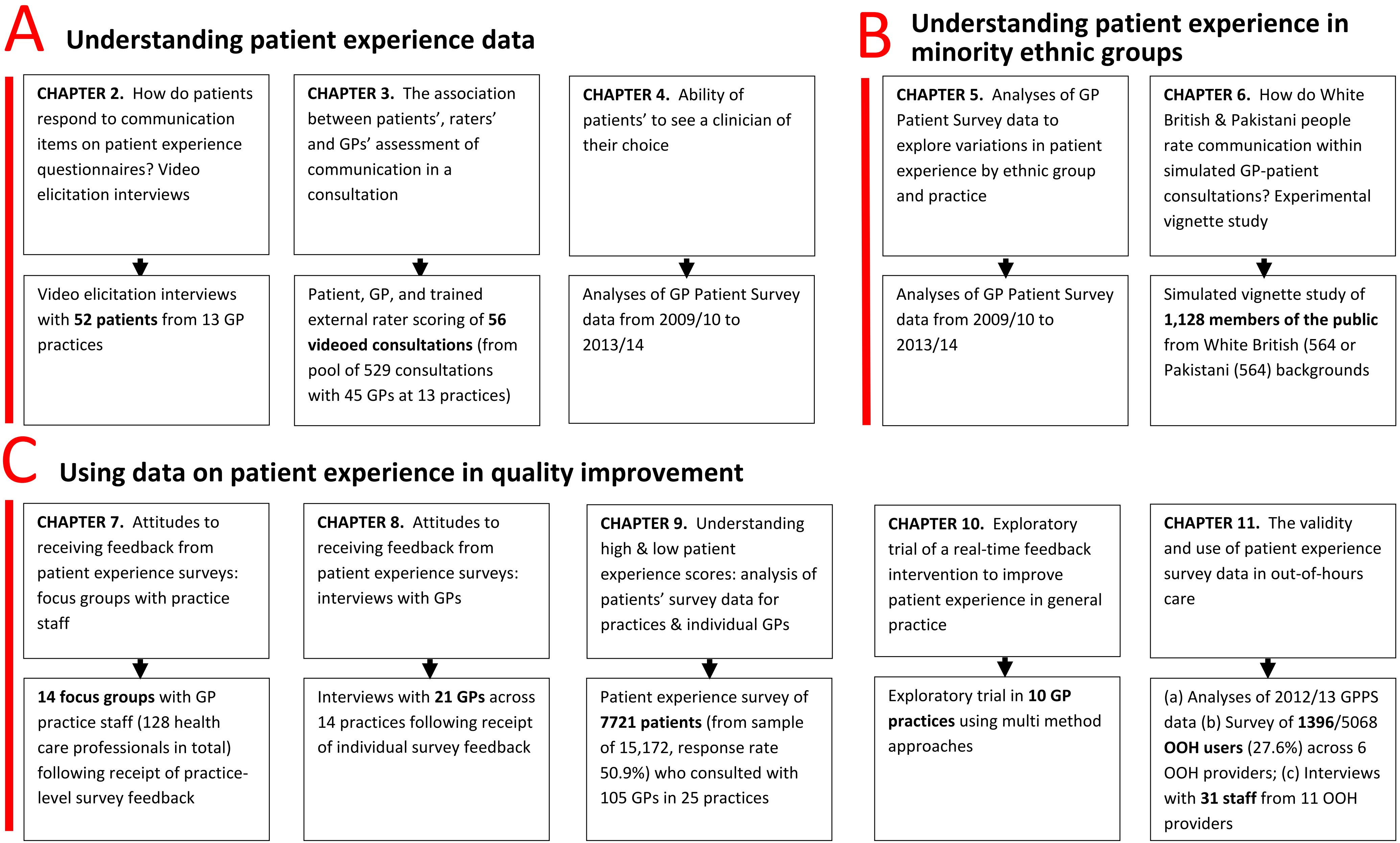
1. To understand how general practices respond to low patient survey scores, testing a range of approaches that could be used to improve patients' experience of care.
2. To estimate the extent to which aggregation of scores to practice level in the national study masks differences between individual doctors.
3. To investigate how patients' ratings on questions in the GP Patient Survey relate to actual behaviour by GPs in consultations.
4. To understand better patients' responses to questions on communication and seeing a doctor of their choice.
5. To understand the reasons why minority ethnic groups, especially South Asians, give lower scores on patient surveys compared to the White British population.
6. To carry out an exploratory randomised controlled trial of an intervention to improve patient experience, using tools developed in earlier parts of the programme.
7. To investigate how the results of the GP Patient Survey can be used to improve patients' experience of out-of-hours care.

In presenting our work, we report our research and findings in three major themes: (a) understanding patient experience data; (b) understanding patient experience in minority ethnic groups; and (c) using data on patient experience for quality improvement. These are outlined in brief, below: the relationships between individual studies and the three themes are shown in Figure 1, which also outlines methods and participants.

During the course of the programme, we conducted empirical studies across a number of GP practices and out-of-hours providers. GP practices were initially recruited to take part in a suite of studies (presented in Chapters 7, 8 and 9) in which we conducted a patient experience survey at the level of individual GPs, gave feedback from this survey to both the practice and the individual doctors (chapter 9) and, for some practices, conducted focus groups with practice staff and interviews with GPs. Sampling was initially designed around the survey study: practices were sampled on the basis of location (two study areas, the South West and North London/East of England, covered both urban and rural settings), performance on the GP Patient Survey, practice size and area-level deprivation**.** Once the survey was completed, a number of practices were purposively sampled to take part in focus groups with staff (Chapter 7) and interviews with GPs (Chapter 8), and additional filming of consultations (Chapter 3). Out-of-hours providers were recruited from across England: we worked with up to 11 providers in varying workstreams (Chapter 11). We additionally completed multiple analyses of GP Patient Survey data (Chapters 4, 5 and 11) and, for an experimental vignette study, collected data from members of the general public (Chapter 6).

1. Understanding patient experience data

In this theme, we explored the meaning of data gathered through patient experience surveys by video recording (with consent) a large number of GP-patient consultations. Patients and GPs completed a questionnaire evaluating the quality of communication within the consultation, and trained external raters (all GPs) also scored a small number of filmed consultations for their quality. We additionally interviewed a sample of patients who consented to have their consultations filmed, reviewing their recorded consultation with them whilst talking through the options they chose on the questionnaire about their experiences. This theme relates to aims 3 and 4. In addition, we conducted analyses of GP Patient Survey data to explore variations in patient experience in patients whose contact is with a nurse rather than a GP. This is additional to the original aims of the programme.



A **Understanding patient experience data**

CHAPTER 2. How do patients respond to communication items on patient experience questionnaires? Video elicitation interviews

Video elicitation interviews with **52 patients** from 13 GP practices

CHAPTER 3. The association between patients', raters' and GPs' assessment of communication in a consultation

Patient, GP, and trained external rater scoring of **56 videoed consultations** (from pool of 529 consultations with 45 GPs at 13 practices)

CHAPTER 4. Ability of patients' to see a clinician of their choice

Analyses of GP Patient Survey data from 2009/10 to 2013/14

B **Understanding patient experience in minority ethnic groups**

CHAPTER 5. Analyses of GP Patient Survey data to explore variations in patient experience by ethnic group and practice

Analyses of GP Patient Survey data from 2009/10 to 2013/14

CHAPTER 6. How do White British & Pakistani people rate communication within simulated GP-patient consultations? Experimental vignette study

Simulated vignette study of **1,128 members of the public** from White British (564 or Pakistani (564) backgrounds

C **Using data on patient experience in quality improvement**

CHAPTER 7. Attitudes to receiving feedback from patient experience surveys: focus groups with practice staff

14 focus groups with GP practice staff (128 health care professionals in total) following receipt of practice-level survey feedback

CHAPTER 8. Attitudes to receiving feedback from patient experience surveys: interviews with GPs

Interviews with **21 GPs** across 14 practices following receipt of individual survey feedback

CHAPTER 9. Understanding high & low patient experience scores: analysis of patients' survey data for practices & individual GPs

Patient experience survey of **7721 patients** (from sample of 15,172, response rate 50.9%) who consulted with 105 GPs in 25 practices

CHAPTER 10. Exploratory trial of a real-time feedback intervention to improve patient experience in general practice

Exploratory trial in **10 GP practices** using multi method approaches

CHAPTER 11. The validity and use of patient experience survey data in out-of-hours care

(a) Analyses of 2012/13 GPPS data (b) Survey of **1396**/5068 **OOH users** (27.6%) across 6 OOH providers; (c) Interviews with **31 staff** from 11 OOH providers

Figure 1. Improve programme themes and studies contained within workstreams

2. Understanding patient experience in minority ethnic groups

Here, we conducted a number of studies to explore why South Asian groups often have lower patient experience scores compared to White British patients in national surveys, and provide more robust evidence of the drivers of this variation. These included a series of analyses of GP Patient Survey data, and an experimental vignette study in which we showed simulated GP-patient consultations to White British and Pakistani respondents. This theme relates to aim 5.

3. Using data on patient experience for quality improvement

In trying to understand how patient experience data is currently used, and how it could be used, we carried out a wide range of studies. We completed a large scale survey of patients at 25 GP practices, and carried out focus groups with practice staff and interviews with GPs. We conducted similar research in out-of-hours services. Finally, we looked at a different way of collecting patient experience data, using “real time feedback” kiosks in GP practices. This theme relates to aims 1, 2, 6 and 7.

Patient and Public Involvement

Our programme of work was supported by two study advisory groups: one, based in Cambridge, which provided support across all streams of work except the out-of-hours research, and one, based in Exeter, convened specifically to provide input to the out-of-hours workstreams. In this section, we briefly outline the formation and working of these groups over the course of the programme.

Formation and composition of the main study advisory group

In the original application for the IMPROVE programme, we set out our plans to establish an advisory group composed of 50% lay and 50% professional members, to provide continuing advice and input throughout the course of the programme. We envisaged that this group would provide advice on the design of all strands of work, assist with the production of study materials, and work with us on the interpretation of data. At the start of the study, we therefore set out to invite a mix of lay people registered with a GP practice, GPs, and practice managers to join the group.

We worked with the PPI Co-ordinator of the West Anglia Comprehensive Local Research Network (CLRN) to identify potential lay members with an interest in patient experience and primary care research. Potential patient representatives were provided with guidance outlining what was involved in advisory group membership, and were informed that any costs incurred in preparing for or attending advisory group meetings would be reimbursed. Four lay members were recruited via this route. Additionally, we recruited one local GP, from a practice with a large minority ethnic patient list, to the advisory group. Despite a number of attempts to recruit an additional GP and two practice managers to the group, to ensure we had input from all key stakeholders in the research, we were unable to do so. In spite of offering reimbursement to practices (for example, we paid for a locum to enable our one GP member to attend advisory group meetings), GPs and practice managers were reluctant to commit to provide input into a research study over a number of years. We therefore proceeded with input from one GP only.

As a large focus of our work was on patient experience in minority ethnic communities, and particularly South Asian communities, we had additionally planned at the start of the study to recruit two additional lay members from a minority ethnic background to join the advisory group and provide specific advice on the development of our work in this area. In the event, this proved very difficult to achieve, and we were unable to locate suitable representatives willing to sit on a formal group. We therefore considered alternative approaches to ensuring we had input from these communities as we developed our study ideas and materials. As a result, we recruited a part-time researcher, Hena Wali Haque, and a senior advisor, Prof Cathy Lloyd, with specific expertise in and knowledge of research with minority ethnic groups. Hena liaised early on in our work with community groups representing Pakistani and Bengali communities, and provided input on study materials and design. Whilst we would have preferred to have had such representation directly on our study steering group, through this route we were able to benefit from guidance on the most appropriate and effective approach to our research in this area.

We drew on guidance from INVOLVE to develop policies and documentation relating to the involvement of our advisory group members. These included details of payment for particular activities, reimbursement, confidentiality, and data security. Group members completed a checklist to indicate what they were willing to assist with during the course of their involvement (for example, reviewing different types of documents or attending meetings).

Formation and composition of the out-of-hours study advisory group

A stakeholder advisory group was convened specifically to provide guidance throughout the out-of-hours research. This comprised three members from out-of-hours service providers, two academics with a particular expertise in this area, and one lay representative. We had originally aimed to recruit two out-of-hours service users through local service providers, with assistance and guidance from local PPI groups: however, despite significant efforts, it proved difficult to recruit service users with relevant, lived experience. Our experiences were echoed by out-of-hours service providers, who noted that the relatively infrequent contacts people made with out-of-hours services may in part drive the difficulties in recruiting service users to sit on advisory groups such as ours.

Activities of the main study advisory group

We set out to convene a face-to-face meeting of the main programme advisory group once a year throughout the course of the research. The first meeting took place in Cambridge in October 2011, and the fifth and final meeting took place in March 2015. At these meetings, group members reviewed and suggested changes to study design, reviewed progress, discussed challenges, and reflected on findings and interpretation. Particularly crucial input came, for example, in designing our approach to the recruitment of patients to our workstream involving the video recording of GP-patient consultations, and in reflecting on the findings of our video elicitation interviews with patients. To keep group members up-to-date with progress and the research team, we sent out study newsletters on a roughly quarterly basis, with thirteen being sent over the course of the programme.

Informal contact with group members via email and letter continued throughout the rest of the year outside of the more structured meetings. One advisory group member, for example, was instrumental in organising a pilot focus group to reflect on study questionnaires. Additionally, all study materials aimed at patients or GPs (information sheets, consent forms, and questionnaires) were reviewed and commented on by advisory group members, and members were sent a summary of all findings and our conclusions to reflect on.

Activities of the out-of-hours study advisory group

Our out-of-hours advisory group, based in Exeter, had a more specific remit in guiding our research in this area. The group met initially to review study methods and procedures in light of the findings of preliminary piloting and testing of methods (see Chapter 11, Workstream

2), and to comment on topic guides supporting interviewing in Workstream 3. However, due to the logistical challenges of organising face-to-face meetings around staff availability, after an initial face-to-face meeting we communicated with the advisory group via email and telephone.

SECTION A

Understanding patient experience data

Chapter 2. How do patients respond to communication items on patient experience questionnaires? Video elicitation interviews with patients

Abstract

Background

Patient feedback instruments used in national survey programmes are robustly tested and evaluated, yet there remains a paucity of evidence on the drivers of a patient's choice of response option. The objective of this study was to understand how patients' responses to a questionnaire relate to their actual experience of a consultation with a GP, focussing on both implicit and explicit processes which respondents use to answer survey items.

Methods

We video recorded GP-patient consultations at 13 practices. Immediately following the consultation, patients were asked to complete a questionnaire about the GPs' communication skills. We purposively approached a sample of these patients to take part in a video elicitation interview (n=52), in which they were shown the video of their consultation and asked to reflect on their completion of the questionnaire.

Results

Whilst participants were able to raise concerns about doctors' behaviours within the interview, they were reluctant to do so in their questionnaire responses. We identified three important drivers of this mismatch: i) the patients' relationship with the GP, ii) the patients' expectations of the consultation, and iii) perceived power asymmetries between patients and doctors.

Conclusions

Patients were inhibited in providing feedback to GPs through use of questionnaires, with patients struggling to transform their experiences into a representative quantitative evaluation of GP performance. Our results suggest that patient surveys, as currently used, may be limited tools for enabling patients to feed back their views about consultations.

Introduction and rationale for the study

The overall purpose of patient surveys in primary care, such as the national GP Patient Survey, is to improve patient experience by feeding back patients' evaluations to GPs and to the public. This process makes an important assumption, which is that the behaviours which doctors need to change are accurately assessed by responses given in patient experience questionnaires. For questionnaire items which relate to doctor patient communication, the evidence that the items reflect doctors' behaviour rests on their face validity and the cognitive testing which has already been carried out. Face validity is often taken as sufficient. However, further understanding of questionnaire completion is needed before helpful advice can be given to GPs. For example, if more is understood about the nature of poor consultations identified by patients, better support and advice can be provided to GPs to improve the quality of their consultations.

Previous studies have examined the process of patient questionnaire completion in specialist clinics: [89, 90] these highlighted that patients may struggle to accurately represent their experiences of a consultation on standard survey instruments. Further, concerns have been raised about the perceived requirement for patients to assess health care from a 'consumerist' perspective. [91, 92]

To date, little is known about the ways in which questionnaire responses relate to patient experience within primary care and, specifically, their perceptions of communication within GP consultations. The aim of this study was to understand, through the use of video elicitation interviews, how patients' responses to a questionnaire relate to their experience of a consultation with a GP.

Changes to study methods from original protocol

The aim of this workstream, as stated in the original protocol, was:

"To understand better patients' responses to questions on communication and seeing a doctor of their choice (aim 4)."

In our application, we set out plans to address this by conducting interviews with 40 patients: 20 to be from a White British background, and 20 to be from an Asian background. Interviews with minority ethnic participants were designed to contribute to our understanding of variations in patient experience of care in these groups, complementing our analyses of GP Patient Survey data and our experimental vignette study (Chapters 5 and 6). We envisaged all interviews drawing on psychological approaches to cognitive interviewing, focussing on (i) comprehension of the question (ii) recall and assessment (iii) decision processes and (iv) response processes.

We have expanded on our original design in several important ways. Firstly, following our application, literature on the use of video elicitation interviews to stimulate recall and reflection on a medical encounter were published, and to us appeared to be of direct utility for the aims of this study. Video elicitation approaches, outlined in full below, use a series of detailed and specific prompts to enable participants to "relive, recall and reflect" on their recent medical consultation. [93] We therefore adopted this approach in preference to that of cognitive interviewing.

Secondly, following discussions with practices, we were concerned that a "one size fits all" approach to recruiting patients from both White British and South Asian backgrounds to the study was unlikely to be sufficiently sensitive and robust. We therefore made the decision to conduct the South Asian interviews as a standalone study, recruiting three additional practices with a particularly large proportion of South Asian patients on their lists and using dedicated researchers fluent in South Asian languages, together with appropriate study materials. This resulted in 23 interviews specifically with patients from a Pakistani background, conducted in the language of their choice. Our analyses of these interviews identified broadly similar concerns between our South Asian sample and that of the main study, and we report these briefly within this chapter.

Finally, we expanded our original sample size of 20 interviews with White British patients to over 50, from a variety of backgrounds (but all fluent in English). Video elicitation interviews are challenging to conduct well, and we felt it was important to enable the research team to build up sufficient confidence and expertise to generate rich data, as well as to reach a more diverse patient population. This chapter focusses in the main on interviews with the English speaking population (n=52).

Methods

This strand of work was conducted alongside the quantitative study outlined in Chapter 3. Recruitment of practices, GPs and patients are thus the same for both: the work outlined in this chapter focusses on subsequent interviews with patients who gave consent for their consultations to be video recorded. The Improve study advisory group made important contributions to study design, particularly our approach to recruiting patients and the use of both a “brief” and a “full” study information sheet, and reflected on our analysis and findings.

Recruitment of GP practices

The study was conducted in general practices in two broad geographic areas (Devon, Cornwall, Bristol, and Somerset; and Cambridgeshire, Bedford, Luton and North London). Practices were eligible if they (a) had more than one GP working a minimum of four sessions a week in direct clinical contact with patients, and (b) had low scores on GP-patient communication items used in the national GP Patient Survey (defined as practices below the lower quartile for mean communication score in the 2009/10 survey, adjusted for patient case-mix (age, gender, ethnicity, self-rated health and deprivation [94]). Low scoring practices were chosen to maximise the chance of consultations within the practice being given low patient ratings for communication: nationally, 94% of patients score all questions addressing doctor communication within consultations as good or very good in the GP Patient Survey. Some, but not all, of these practices had previously participated in our individual-GP level patient experience survey (see Chapter 9 for details).

Recruitment of patients and recording of consultations

Video recording of GP-patient consultations took place for one or two GPs at a time within each participating practice. A member of the research team approached adult patients on their arrival in the practice to introduce the study. Patients were given a summary of the study within a brief information sheet, as well as a detailed full information sheet, and a consent form. A member of the research team discussed these with each patient and sought consent to video record their consultation. Video cameras, set up in participating GPs’ consulting rooms, were controlled by the GP: physical examinations took place behind a screen and were thus not captured on camera. Data collection ceased when we reached our required number of video-recorded consultations which patients judged to be less than good for communication, as required for the quantitative analysis described in Chapter 3. All videos were stored on an encrypted secure server accessible only to members of the core research team. Recordings were made available to GPs for the purposes of continuing professional development. Immediately after the consultation patients were asked to complete a short questionnaire. This contained items relating to GP communication adapted from the General Practice Patient survey (Box 1) alongside participant information including age, ethnicity and health status.

Thinking about the consultation which took place today

How good was the doctor at each of the following?

Please put an × in one box for each row

	Very good	**Good**	**Neither good nor poor**	**Poor**	**Very poor**	**Doesn't apply ***
Giving you enough time	☐	☐	☐	☐	☐	☐
Asking about your symptoms	☐	☐	☐	☐	☐	☐
Listening to you	☐	☐	☐	☐	☐	☐
Explaining tests and treatments	☐	☐	☐	☐	☐	☐
Involving you in decisions about your care	☐	☐	☐	☐	☐	☐
Treating you with care and concern	☐	☐	☐	☐	☐	☐
Taking your problems seriously	☐	☐	☐	☐	☐	☐

Box 1. GP-patient communication items used in the patient experience survey

Video elicitation interviews and analysis

The patient questionnaire contained a tick box asking patients if they were willing to participate in a face-to-face interview about their experience of the consultation. We subsequently contacted (via telephone or email) those patients who expressed an interest in taking part. We aimed to interview at least one patient per participating GP. When more than one patient expressed interest, we used a maximum variation sampling approach to reflect a mix of patient characteristics and questionnaire responses. Prior to the commencement of the study, we were particularly interested in interviewing patients who had given at least one response of 'poor' or 'very poor' in relation to a doctor's communication skills.

We conducted video elicitation interviews with all participants (n=52). In these, participants were shown a recording of their consultation with the GP and asked specific questions relating to the consultation and their questionnaire responses (See Box 2 for full details). The video elicitation technique is an established interview method which allows in-depth probing of experience during the interview by enabling participants to 'relive, recall and reflect' on their recent consultation. [93]

Video elicitation interviews

Data generation focussed particularly on participants' recall of and reflection on the consultation, and how this was expressed in their choice of responses on the questionnaire immediately post-consultation. In each interview, the video of the consultation was used to encourage more accurate recall of specific events during the interaction. Our approach did not aim to establish the facts of what occurred, but rather explored the meaning to patients of actions that were performed in the consultation. The interview guide used was semi-structured; however, we maintained a tight focus on specific moments and events captured in the recording.

Participants were asked some brief introductory questions about whether they had previously consulted with this doctor, and whether the problem they were consulting about was new or ongoing. Patients were then shown their consultation on the researchers' laptop. They were encouraged to reflect as they watched the recording. Patients were also given their questionnaire responses and invited to talk through them. The recorded consultation was used as a prompt, enabling further in-depth discussion of their experiences in the consultation and their responses to the survey questions. Patients were also asked to identify behaviours in the consultation that they considered as contributing to their question responses and which could be changed to improve consulting performance.

Box 2. Video elicitation interview approach

Interviewers watched the consultation usually on at least two occasions before the interview and identified particular points at which they wished to stop the recording, or where they wished to use prompts specific to the consultation content or the respondents' answers on the questionnaire. During the interview the video recording of the consultation was shown to the participant, usually on two occasions. Participants were encouraged to stop the recording at any point to discuss a particular element of the consultation with the interviewer. The interviewer also stopped the recording as appropriate in response to a request from the participant, something said by the participant or the interviewer's own prepared notes.

Analysis followed the principles outlined by Lofland and Lofland. [95] These form a series of reflexive steps through which data are generated, coded, and re-coded, making particular use of memos to aid analytical thinking. Data analysis took place in two stages. The first stage occurred during data collection. A coding frame was devised from the topic guide, previous literature and early interviews. Each interviewer (JN, NL, and AD) coded their own interviews in NVivo v.10 software (QSR International Ltd, 2012). A number of analysis meetings were convened in which the interviewers and other members of the project team (JB, NE and JBe) discussed the data and themes. To ensure familiarisation with all the data the lead author (JN) listened to all interviews and read all the transcripts. The coding frame was refined in response to discussions and as analysis progressed.

Approval for the study was obtained from the NRES Committee East of England – Hertfordshire on 11 October 2011 (ref: 11/EE/0353).

Results

Participant recruitment

Consultations were videoed with 45 participating GPs from 13 general practices. During the period of data collection a total of 908 patients had face-to-face consultations with participating doctors. Of these, 167 (18.4%) were ineligible (mostly children), and 529 completed a questionnaire (71.4% response rate) (see Figure 2 for details).

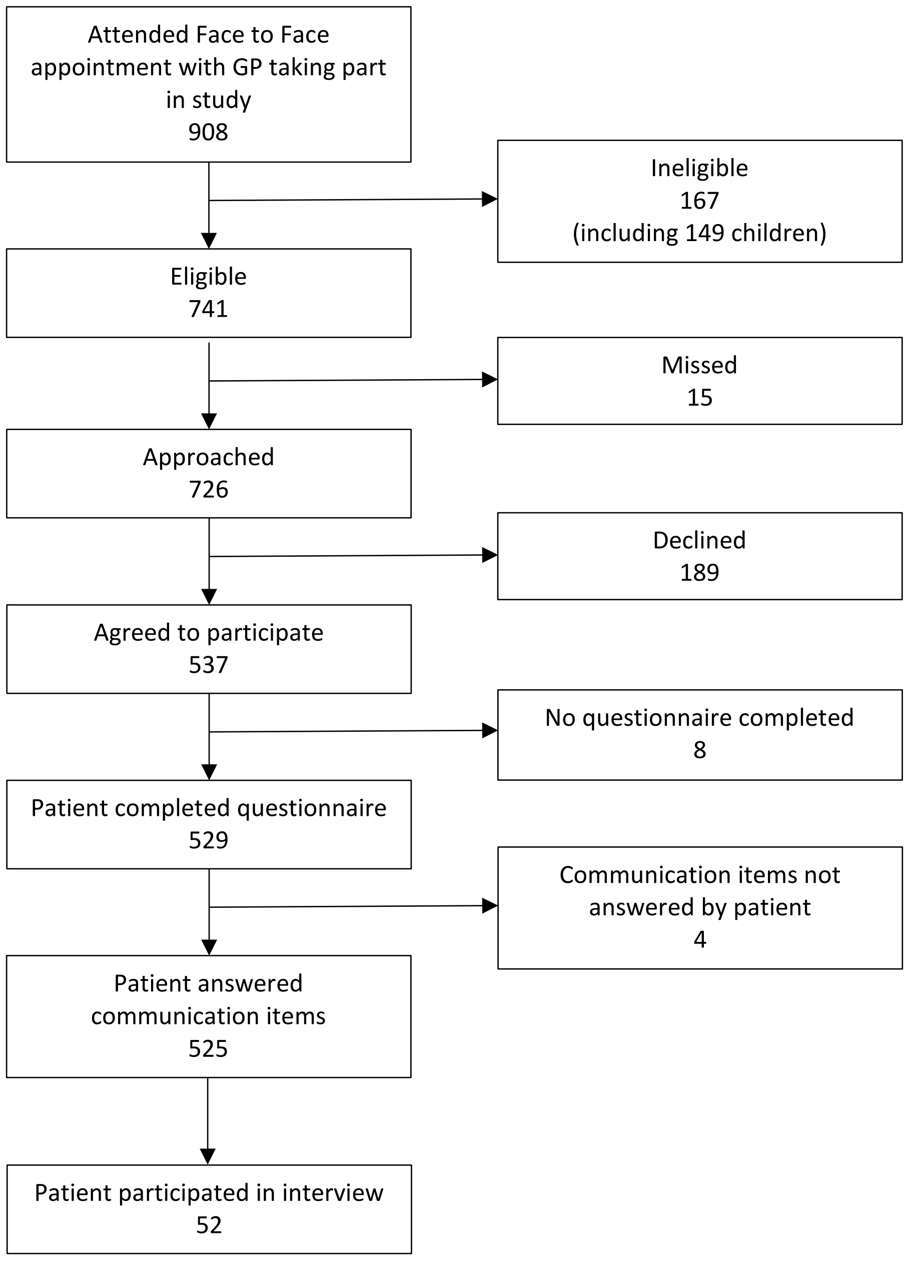


Figure 2. Flow of patients through the video elicitation interview recruitment process

Video elicitation interviews

A sample of patients whose consultation was video recorded participated in a video elicitation interview. In total interviews were conducted with 52 patients (35 women, 17 men) who had consulted with 34 different doctors across 12 GP surgeries in rural, urban, and inner-city areas in the South West and East of England.

Interviews took place between August 2012 and July 2014. Interviews were conducted within a maximum of four weeks from the recorded consultation in a location chosen by the participant (Table 1). It was the researchers' preference that interviews were not usually conducted in the general practice surgery in case it inhibited patients in their narrative. However, a few participants specifically requested that interviews be held at the GP surgery.

Table 1. Location of video elicitation interviews

Location	Number of interviews
Participants' home	44
GP surgery	6
Other location (chosen by participant)	2

All interviews were conducted in English, and lasted between 26 and 97 minutes (average 58 minutes). Participants were aged between 19 and 96 years of age. 22 participants (42%) were over 64 years of age, and 30 participants (58%) were aged between 19 and 64 years. Participants consulted for a range of conditions, some chronic and some minor illnesses. Respondents names used in this section have been changed and are not the real names of participants.

Questionnaire completion

In interviews participants were well disposed towards the process of questionnaire completion and generally keen to contribute their views. Most participants described completing the questionnaire with relative ease and as a simple task. Despite this willingness to contribute, there was little variety in questionnaire responses: the majority of participants reported care to be good or very good across all seven communication items on the questionnaire. Indeed, no respondents in our interview sample chose to score 'poor' or 'very poor’, despite our original aspiration to focus in particular on patients who expressed dissatisfaction with their care. Twelve respondents did, however, use the ‘neither good nor poor’ option in at least one domain, although five of these also scored ‘very good’ on at least one other domain. As a result, in our small sample we had a lower proportion of scores in which every domain of GP communication was judged to be ‘good’ or ‘very good’ compared to the national GP Patient Survey sample (77% in our sample vs 94% nationally). Thus, despite the lack of ‘poor’ responses, we were able to explore patients’ responses in those who had expressed more dissatisfaction than average.

Disconnect between the ‘tick and the talk’

Whilst scores on the questionnaire were largely positive, some narratives in the interviews were more critical of aspects of GP communication. We outline three types of narrative relating to the relationship between questionnaire responses and further reflection on the consultation experience expressed within the interview.

1. Re-watching the consultation endorsed positive questionnaire scores

For some participants, their reflection on the consultation within the video elicitation interview led to a repeated endorsement of the questionnaire responses they had given, and thus their narrative account was consistent with their previous evaluation of care. In all cases these responses were positive. Participants had been pleased with the quality of the consultation at the time of completion of the questionnaire. On re-watching the consultation this view was endorsed and in some cases further strengthened. Some respondents pointed to elements in the videoed consultation which had impressed them:

> *… his (GP’s) movements, his mannerisms…I’m asking the question, he didn’t exactly ignore me, he says no, that’s for gout. He actually explained it…And he’s still doing some work…So he’s not stopped and put all his attention on me, because if you stop doing that you probably forget what you’re doing here, so he’s done both. He’s answered my question and he’s also continued working, and that’s a good thing for me.*
>
> *Colin (53151034)*

2. High quantitative scores were followed by some criticism in interviews

Some participants scored the consultation highly on the questionnaire, yet the subsequent interview was peppered with tones of criticism about aspects of the consultation.

Criticism in the interview was often subtle, with participants often seemingly unaware of the discrepancy between their narrative and questionnaire responses. Even though they spoke of their consultations in a tone which was not particularly positive, participants remained loyal to the positive scoring they had applied on the questionnaire immediately following the consultation:

> *I gave it 'good' because…well she was listening to me, but I guess most of the time she was the one talking rather than listened to what I was saying …Not in a negative way, like completely, but I feel she didn't really give me proper time to properly explain myself a little more. ..giving me a little bit more time, to explain my symptoms*
>
> *Steven (60121017)*

3. Participant re-appraises the consultation during the interview

A small group of participants, who had scored their GP highly on the questionnaire, underwent a process of re-appraisal of the consultation during the video elicitation interview. They voiced criticism of the doctor's behaviour and proceeded to review their original score. Through the process of re-watching, participants spontaneously identified more negative aspects of the consultation which they had not been aware of previously:

> *Jack: I suppose you're proving to me that I marked that wrong (taps questionnaire) [laugh]….Yeah, but he (GP) did, he did, he was concentrating on my leg and not worrying about the fact that the tablets were upsetting me.*
>
> *I: Mm. And how did you feel?*
>
> *Jack's wife: Well, I felt the same thing. He, sort of, ignored the fact that he'd got all these side effects and all that.*
>
> *Jack (55161002)*

Emma had scored elements of her consultation as 'very good' on the questionnaire:

> *…now I'm thinking, well no, he didn't really sort of ask about symptoms or think, y'know, so perhaps not so good. Listening – yeah he listened but didn't pick up on things, like you say, like the cough, he didn't sort of pick up on erm, little things*
>
> *Emma (27131004)*

On occasion there was a dramatic shift in point of view when the consultation was re-shown. During the re-watching of the interview, Martha began to critique more aspects of the consultation, such as the doctor's lack of explanations and unexpected examination;

> *I remember him just like, because he, because it's quite rushed,… you, er, can't, you don't, I don't know, you're just, it's just like, er, er, and then, fine, I don't know, I suppose I remember thinking why is he taking my temperature, and then just seeing how it must be okay, erm, I, I definitely remember him when he was just doing that with my, feeling my neck [slight pause] wondering what he was doing. [laughs] I just remember thinking, this is a bit weird, like why is this connected to my ear.*
>
> *Martha (62111010)*

In a number of the interviews, therefore, there was a mismatch between the subsequent account and previous responses to questions. At times participants were happy to critique an experience during the interview, sometimes at great length, yet they had been reluctant to do so on the questionnaire. Participants were able to explain in great detail elements of the consultation which they experienced to be negative, yet when asked to complete the questionnaire on that basis, they still scored the doctor as 'good'. The use of the video elicitation method identified the possibility that other factors fed into the choice of response options on the questionnaire, aside from the doctor's behaviour in the consultation.

There was therefore, on a number of occasions, a disconnect between the 'tick' and the 'talk': differences between the narrative given in the interview and the responses recorded previously in the questionnaires. Whilst participants were able to raise concerns about doctors' behaviours within the interview, at times they appeared reluctant to do so in their questionnaire responses.

Factors which influence patients' reluctance to criticise on the questionnaire

This reluctance to record negative responses on the questionnaire leads to the question as to why patients were reluctant to do so, given the negative views often apparent in their narratives. We therefore sought to further understand this phenomenon. We identified three key factors which appeared to influence patients' reluctance to criticise doctors' communication skills within the questionnaire:

1. the patients' relationship with the GP
2. the patients' expectations of the consultation
3. perceived power asymmetries between patients and doctors.

The following sections of this chapter will examine each of these explanations in turn.

1. The patients' relationship with the GP

Participants often spoke about the significance of the GP or the surgery in their lives. This affiliation was sometimes with the practice, even if the doctors had changed over the years. Some elderly patients interviewed had been with their surgery most of their lives, and a number of participants expressed loyalty to a practice even if they did not often visit. Gratitude for the wider health service and NHS provision were views also commonly expressed.

In commenting on relationships with individual GPs, care given previously to the participant or to their family and friends were often praised:

> *…but I mean I've known him for - I mean he actually phoned when my mum died, you know. So that was nice of him, you know.*
>
> *Janice (67131043)*

Some participants spoke particularly of their relationship with the doctor. In some cases this was the notion of 'getting on with' the doctor and liking them as a person. For some there were specific interests which were shared, such as an interest in sport or knowledge of the GP's family:

> *Oh yes, yes [laughter] they go to my church as well you see, so in that sense the relationship I had with Fiona and Paul is very much the old fashioned family doctor, where you know them.*

Alan (19144016)

Some participants used the term 'friend' to describe their relationship with their GP, often going on to explain the relationship was different to a friendship;

> *And as I said, y'know, he isn't a friend but you feel as if you are seeing a friend.*

Bob (25111005)

> *I can see now the relationship. I have to be careful, you know, when I said to somebody one day well, you and I have a very good relationship, and I thought oh no, that's not the right word*

Janet (53181024)

This was the case for participants who had not previously seen the doctor they consulted with, as well as those who had a long relationship with them. However, for some it was difficult to create a relationship with the GP, which could lead to challenges in building rapport with the doctor.

Most respondents articulated that they were responding to the questionnaire based on the recorded consultation at hand. However, in explaining their scores participants would often reflect on previous consultations with the GP. Consequently, it appeared challenging for them to differentiate between this particular consultation and their wider relationship with the GP in evaluating communication skills. This loyalty and closeness with the GP at times inhibited patients from negative survey responses.

2. The patients' expectations of the consultation

A variety of patient expectations influenced the scoring process. Firstly, we identified expectations which related to the communication skills of the GP, based either on previous experience of consulting with that same GP, or in comparison with experiences of consultations with other GPs. Expectations were important, and participants used this relational knowledge of other GPs to compare care received in the consultation with experience of previous consultations as they reflected within the interview. Some participants compared the care received to high quality care from other doctors. More commonly, however, participants compared a GP's behaviour to poorer care they had received. In particular, patients appeared to bench mark GPs' skills based on their experiences with other GPs:

> *And he's (GP), he's not as clipped as Dr Williams, but he can be sometimes a bit clipped in the way he speaks to you.*
>
> *Dave (67131012)*

Narratives covered both positive and negative expectations of care from a particular GP, so, for example, if the participant had high expectations and these were not met they were disappointed. Conversely, if a participant had a poor expectation of a GP but the consultation was better than they had expected, they might score the GP more positively, even if overall their experience of the consultation was poor.

Secondly, expectations relating to the outcome of consultations were often referred to in justifying questionnaire scores. Participants tended to rate consultations more positively on the questionnaire if the outcome was what they desired, for example if they wanted a particular type of medication, a referral, or reassurance about their medical problem:

> *I worry that, like, yeah, that he's just going to be really dismissive. So the fact that he gave me medicine meant that it was higher than I expect …like it was better than I expected it to be, em, but perhaps by the more standards it wasn't amazing.*
>
> *Martha (62111010)*

3. Perceived power asymmetries between patients and doctors

Descriptions of power asymmetries were prevalent in many of the accounts. Participants were often reluctant to criticise GPs, fearing repercussions for the GP. One respondent, who shared a story of poor experience as a patient in her GP practice, had scored the GP 'neither good nor poor' on some elements. When asked why she had not used the 'poor' or 'very poor' options on the questionnaire, she replied:

> *…you don't want to get anybody into trouble, you know, but you do wish they did behave a little better, you know, treated you a little better, you know, in their response to you.*
>
> *Esther (53131010)*

At times there was an associated dependency on the GP expressed by participants, with a corresponding view that they could not be critical for fear of compromising the relationship.

Participants often spoke of the trust they placed in the doctor. For some, the doctor had a status they were in awe of. Although Amanda felt this relationship had changed over time, the doctor was still revered:

> *A: And I think the gap is not as wide as it used to be between doctors and patients is it? I mean when I was a young girl the doctor was even more of a god, whereas now it's less, it's definitely getting lesser, yea definitely*
>
> *I: And in terms of the GPs at the surgery, I know you mentioned about GPs being on a pedestal, how do you feel the GPs are there?*
>
> *A: Oh I think they all are, yea definitely yea. But years ago perhaps it was a six foot pedestal whereas now it's probably a couple of foot [laughs]*
>
> *Amanda (24111009)*

In some ways participants had an inability to critique the doctor, or at least reluctance to do so. For example, some participants seemed to feel they lacked the authority to judge the doctor's communication skills.

Allowances were often made by participants for elements of the GP's behaviour which they did not like, such as the GP looking at the screen a lot during the consultation:

> *She (GP) was reading so and I mean there's an awful lot on there [laugh] there's loads on that screen, bless her, so she's probably thinking oh my God, how many [laugh] but no I don't take a lot of notice to be honest*
>
> *Sue (24155004)*

Respondents often commented on how busy the doctor was that day, or how much they had to do. Participants were on a number of occasions dismissive of other patients and the unreasonable requests they made of GPs.

During interviews, participants could be critical of their own behaviour, taking responsibility themselves for poor communication in the consultation;

> *I: What makes you say that you weren't a very good patient?*
>
> *P: Because I was spending too much time…I wasn't giving out information as clearly as I should do, and, you know, I had gone in with an agenda.*
>
> *Philip (60111001)*

Additional interviews with South Asian respondents

Alongside the interviews conducted as part of the main study, we set out to recruit participants from a South Asian background, to explore in particular their experiences of GP care and the factors which influenced their choices of response options on patient experience questionnaires. We followed exactly the same procedures as for our main study sample, but worked in three practices (in Bedford, Peterborough and Luton) which had a high proportion of Pakistani patients on their practice list. Our researchers for this workstream were fluent in Urdu and Punjabi, and all study documents were available in Urdu. We followed standard procedures for forward and backward translation of these documents.

We conducted 23 video elicitation interviews with respondents who self-identified as Pakistani (eighteen male and five female; aged between 18 and 74 years). Study transcripts were analysed separately to consider the determinants of patient experience and questionnaire response tendencies in this patient group, following the same approach as outlined above for the main study sample. Only three participants chose “neither good nor poor” as at least one response option for the communication items, with none choosing poor or very poor.

In line with our main study sample, respondents were broadly positive about the GP’s communication skills when asked directly. Respondents were able to identify a number of approaches used by GPs which they rated positively – for example, in explaining tests and treatments:

> *“Yes, so I liked that [referring to the recording], the way she showed me on the...err.. she had like a diagram of a body, she even pointed to…like, like… the nerve and where I’ve got a spasm. So I like a bit of that. So that explains to me more of the situation…So yeah, that part was quite good.”*
>
> *Tahir (66 18 5090)*

And in relation to being given enough time:

> *“Obviously they’re very short of time, and he was obviously still getting prepared to see me, when I came in. So when I started the conversation he was still looking at the screen, but he immediately, once he got through that piece of work, he immediately established eye contact, which again helps, certainly, to put me at ease, and know that somebody’s listening, responding and understanding. So that was good.”*
>
> *Sajid (65 13 5113)*

However, within interview narratives it became evident that they drew on a number of factors external to the immediate encounter with the GP in evaluating their communication competence. In spite of differences in cultural background, we identified the same issues driving evaluations of communication. Firstly, respondents often drew on their relationship with their GP, and others within the practice, in making evaluations of care, rather than the specific events in a consultation. Many participants expressed high regard for their current GPs, and often compared them with GPs from their past to explain why they considered them to be so good:

> *"I've been to other surgeries as well… and they're not really interested, they just want to get you in and out, but this practice itself, the doctor listens to you, gives you a lot of time, yeah."*

Imran (65 13 5110)

Secondly, expectations of the consultation could influence assessments of care. For some Pakistani respondents, experiences of health care abroad and a sense that the NHS provided high quality care meant that, regardless of a consultation experience, they were grateful for and positive about encounters:

> "*In Pakistan, if you have money, it is okay otherwise you are on the road. There are very competent doctors in Pakistan, but you need a lot of money for them. Everyone can't afford that -some can and some can't… It means a lot to me. Big deal for me! Fine. Allah forbid, if I have to go to private, I can't do it, can I? I can't afford it. So we are on NHS account.*

Mohammed (65 13 5085)

Finally, perceived power asymmetries between patients and doctors were often prominent – as one respondent clearly articulated, a core theme in Pakistani patients' narratives of experience was that doctors were perceived as having a sacred position in society:

> *"The profession of a doctor is holy. Life and death is commanded by Allah. He saves, but the doctor is the best means."*

Mohammed (65 13 5085)

Limited English proficiency did, however, have the potential to compound these issues: for example, participants could additionally blame their own difficulties in communicating as contributing to challenges within the consultation:

> *"Our issue is the language, all doctors are good. Why would we question them? They have studied to be good, so for me the language creates hurdles."*

Anam (65 13 5061)

Discussion

Whilst participants commonly showed reluctance to criticise GPs in their survey responses, our video elicitation interviews opened up a more nuanced discussion in which patients voiced a number of criticisms about the communication skills of the GP. Previous studies have identified a reluctance to critique doctors in users of mental health services [90] and patients undergoing elective orthopaedic surgery. [89] Our findings confirm that such reluctance persists for some patients in general practice, a setting in which an ongoing relationship between doctor and patient is more commonplace compared to the secondary healthcare setting.

Medical encounters have long been characterised by an asymmetrical power balance, despite attempts over the last fifteen years of Government policy [96] to embrace a more ‘patient centred’ approach to healthcare provision. Goodyear-Smith and Buetow have urged that, whilst seeking to empower patients, we must be sure not to disempower doctors; they note that power can be beneficial in the consultation. [97] Others have also argued that asymmetry is essential to the success of the medical encounter. Pilnick and Dingwell [98] distinguish between functional and dysfunctional asymmetry in this role, arguing that the former may prove useful in shaping the medical encounter. Within our findings the notion of power asymmetry was evident in many narratives, and this was not necessarily viewed negatively by participants. In fact, patients’ accounts often displayed a respect and at times even reverence towards GPs and the work they do. However, this relationship may make it difficult for patients to be critical when giving feedback to doctors, and they therefore may need more encouragement or permission to report negative aspects of the consultation in questionnaire responses.

The role of expectation was important in our study. Previous work has suggested that non-fulfilment of expectations of care, such as examinations, tests and referrals, can be associated with lower patient satisfaction. [99, 100] Our work suggests that reported experience may be influenced by the meeting of overall expectations, even if the overall standard of the communication was not seen favourably by the participant.

Throughout our interviews, we identified a number of ways in which patients may be inhibited in choosing negative responses on experience questionnaires. Lupton, in her examination of the concept of consumerism in health care, asserts that participants can hold both a ‘consumerist’ and ‘passive patient’ position simultaneously or variously in interactions with doctors. [92] In this study, patients appeared to struggle to inhibit the purely consumerist approach to health care, as Lupton found in her earlier work with patients in Australian general practices. [91] Health may hold vested emotional significance in patients' lives, making it more challenging for them to provide "objective" assessments of care so common in other areas of consumer experience. Coyle's work has also highlighted the personal nature of health care experiences and the threat to personal identity experienced when problems with healthcare provision occurred. [101]

Our study identified a number of contextual factors, including power dynamics, expectations of care, ongoing relationships and previous experience which impact on a patient's choice of response to a questionnaire concerning their experiences. For some patients, this translated into an inhibition to provide negative evaluations of care on a questionnaire, despite being able during interviews to identify a number of concerns about the quality of communication they experienced. Questionnaires, whilst an important tool for gathering patient feedback, may be limited in the information provided in their absolute scores. Our quantitative evaluation of patient assessments of care, reported in Chapter 3, provides more details on how questionnaires may best be used for quality assurance and improvement initiatives. We note that GPs' professional development may benefit from other methods of feedback in addition to patient questionnaires, such as recording and reflecting on their own consultations or having peers watch and discuss their consultations.

Strengths and limitations

Our use of video elicitation methods enabled us to probe in detail the link between a patient's responses on a questionnaire and their experience of the consultation. A number of patients attending participating practices declined to have their consultation recorded. It may be that these patients had particular conditions where they may have been more conscious of their privacy, such as gynaecological or mental health issues. We acknowledge that, as GPs and patients knew that the consultations were being video recorded, this may have altered behaviours. However, as the camera was in the room for most of the day a number of GPs commented on how its presence became normalised during the session. GPs were able to opt in or out of the study, and it may be that doctors who were less confident in their communication skills declined to participate.

The use of the video elicitation method and ability for patients to experience the consultation through re-watching it after the event created a unique environment. Inevitably, the method prompts patients to reflect on a consultation in a novel way. The temporal element of the experience in the consulting room was emphasised and the re-experiencing sometimes led to an altered view of the consultation. For example, the critical self-reflection seen in the data may in part be an artefact of the method in which participants viewed themselves in the consultation during the interview. We also note that the time delay between the consultation and interview may mean that any number of events (for example, a worsening of their condition) may lead to a re-evaluation of the nature of the consultation and a more negative critique, particularly as the patient was (in most cases) further removed from the GP practice and the consultation.

Whilst researchers gently prompted participants regarding their responses to the questionnaire, it may be that some felt the need to give an account of their responses in a socially acceptable way rather than their actual thoughts at the time of interview. For example, they may have preferred to present a rationalised explanation for their responses rather than admit they rushed the questionnaire and did not give consideration when completing it. For interviews with Urdu speaking patients, we used materials translated into Urdu: we did not, in this qualitative work, consider the cultural equivalency of the translated instrument using consensus meetings, and there may be unidentified issues in understanding as a result. However, our bilingual interviewer had the opportunity to draw on shared understanding of concepts during interview, albeit in an ad hoc manner.

Conclusions

Our findings suggest that patients may, on occasion, be inhibited in providing feedback to GPs through a questionnaire. The factors we identify may account for some of the tendency of patients to score consultations highly on questionnaires, with issues including previous experiences, ongoing relationships, and perceived power asymmetries contributing to evaluations of communication skills. Our results suggest that patient surveys, as currently used, may be limited tools for enabling patients to feed back their views about consultations. Doctors whose communication skills are rated ‘very good’ on a patient questionnaire are likely to conclude that no change in their consultation style is required; however, this work suggests that even a rating of ‘very good’ may in fact mask patient reservations about the quality of the encounter.

Chapter 3. The association between patients’, raters’ and GPs’ assessment of communication in a consultation

Abstract

Background

Whilst patient feedback is widely used with the aspiration of quality improvement, the association between patients’, external observers’, and GPs’ own evaluations of communication performance within a consultation remains little explored.

Methods

We video-recorded 529 consultations with 45 GPs in 13 practices. Following the consultation, patients rated the GP’s communication skills, and the GP did likewise. Subsequently, 56 consultations were sampled to include a range of patient scores for communication. Each video was rated by four trained clinical raters using the Global Consultation Rating Scale. The ratings of patients, raters, and GPs were compared.

Results

There was a modest positive correlation between patient ratings and those made by trained raters (rho=0.29, increasing to 0.33 after accounting for measurement error/reliability, p=0.054). Consultations scored highly for communication by trained raters were also scored highly by patients. However, when trained raters judged communication to be of lower quality, mean patient scores ranged from “poor” to “very good”. There was no evidence that GP scores were associated with the scores of trained raters (p=0.721) or with the scores of patients (p=0.854).

Conclusions

Compared to patients, trained rates tended to give more negative scores on communication within consultations. This is consistent with the finding from patient interviews that patients find it difficult to criticise GPs when completing questionnaires. Patient surveys are a useful tool for measuring relative performance of doctors’ communication skills, but absolute scores should be interpreted with caution. Our results also cast doubt on how useful doctors’ own assessment of their own performance is when used as part of reflective practice.

Introduction and rationale for the study

A clear aspiration for the national GP Patient Survey programme is to facilitate changes in overall experience of care by feeding back patients' evaluations both to GPs and to the wider public. Confidence in the instruments used to assess – and potentially rank – performance is therefore essential if they are to make a meaningful contribution to quality assurance and improvement. [102] There has been extensive work on the reliability and validity of patient experience questionnaires. [103-108] However, whilst the face validity of communication items in questionnaires such as the GP Patient Survey has been well studied, evidence is sparser on whether the scores have construct validity: that is, whether behaviours which doctors may need to change are accurately represented by responses given in the questionnaires. For example, do patients reflect specifically only on their experience of communication with the GP in their choice of response options, or are they drawing on wider influences which may be internal or external to the consultation? And how do patients concept of "good" communication relate to professionally agreed norms of "good" communication?

One approach to investigating the construct validity of items is to compare patient evaluations of consultation behaviours with those of external observers. Previous research has explored the relationship between patient and examiner ratings of trainee GP communication skills, and has found either no evidence of an association (in an underpowered study, with a sample size of 19) [109] or weak to moderate association. [110] This workstream aimed to provide more robust evidence of the association between patient assessments of communication skills on items used in national survey programmes, and observer assessment of the performance of practising GPs.

An additional area of concern for quality improvement efforts is that, despite the extensive psychometric testing of patient experience instruments, research shows that doctors often struggle to make sense of and act upon feedback from patient surveys [111] (see Chapter 8 for our work on this). A possible contributory factor in this may be incongruence between self (doctor) and patient assessments of performance. Evidence suggests that doctors tend, in fact, to rate themselves more negatively than patients or peers. [112, 113] Indeed, there is a substantial body of evidence that doctors' perceptions of their own competence are frequently out of kilter with external assessments of the same. [114-116] Of particular concern, however, is that the highest levels of incongruence are found in doctors who are, by external evaluation, the least skilled but the most confident in their abilities. [114, 117]

Previous research has tended to focus on the associations between doctor- and other-assessment at the level of overall performance, rather than performance at the level of a particular consultation. To understand in more detail where discrepancies arise between doctor- and patient- assessments of care, this workstream also considered how GPs' and patient assessments of communication compared at the level of the individual consultation.

Changes to study methods from original protocol

The aim of this workstream, as stated in the original protocol, was:

"To investigate how patients' ratings on questions in the GP Patient Survey relate to actual behaviour by GPs in consultations (aim 3)".

Our application envisaged this workstream taking place as part of our wider patient experience survey, with participants being drawn from patients attending the lowest ranking 15 (out of 25) GP practices (programme aim 2, reported in Chapter 9). This would have entailed asking patients for consent to film their consultation as well as consent to participate in the exit survey, planned to take place face-to-face. However, with the change in survey mode from face-to-face to postal (see Chapter 9 for details), we made the decision to separate this study entirely from the larger scale survey. We thus recruited a sample of low-scoring practices specifically to participate in the filming of consultations, making this a completely stand-alone piece of work.

As this became a fully separate study, we were able to additionally ask GPs to rate, after each video-recorded consultation, their communication performance. This enabled us to undertake additional, originally unplanned, analyses on how GPs' perceptions of their own performance relate to that of the patient or external raters.

Further, in our original application we planned to ask external raters to use the GP Patient Survey communication items to evaluate the quality of communication within a consultation. However, we decided to take a more robust approach to assessment, and thus developed our own instrument to assess communication quality (the Global Consultation Rating Scale), based on a widely used international approach to communicating within a consultation, the Calgary-Cambridge guide to the medical interview.

Finally, our original application had a focus on identifying specific behaviours which may have been associated with patient reported communication scores. In particular, we were interested in identifying which dimensions were of most importance to patients, and thus those which GPs might want to change in order to improve patient experience. However, the study design was powered to detect an overall association between patient and other ratings. The consequence of this is that we were underpowered to differentiate between different doctor behaviours, and this was confirmed in initial analysis. Given this realisation, we have chosen not to present that analysis and to concentrate instead on the overall associations.

Methods

This study took part alongside the video elicitation interviews described in Chapter 2. Briefly, we obtained consent from patients and GPs to video-record face-to-face consultations in participating practices. Full details of our approach to sampling, recruitment and recording of consultations are given in Chapter 2. As already stated, immediately following the consultation, the patient was asked to complete a short questionnaire including a set of seven items taken from the national GP Patient Survey to assess GP-patient communication (Box 3), and basic socio-demographic questions. At the same time, the GP answered the same questions about their own performance. We calculated two doctor-patient communication scores, one from the patient responses and one from the GP responses. In line with previous work, we calculated communication scores by linearly rescaling responses between 0 and 100 and taking the mean of all responses where four or more informative answers were given. [118-120]

Thinking about the consultation which took place today
How good was the doctor at each of the following?
Please put an × in one box for each row

	Very good		Good		Neither good nor poor		Poor		Very poor		Doesn't apply *
Giving you enough time..........................	☐	...	☐	...	☐	...	☐	...	☐	...	☐
Asking about your symptoms	☐	...	☐	...	☐	...	☐	...	☐	...	☐
Listening to you	☐	...	☐	...	☐	...	☐	...	☐	...	☐
Explaining tests and treatments	☐	...	☐	...	☐	...	☐	...	☐	...	☐
Involving you in decisions about your care	☐	...	☐	...	☐	...	☐	...	☐	...	☐
Treating you with care and concern	☐	...	☐	...	☐	...	☐	...	☐	...	☐
Taking your problems seriously	☐	...	☐	...	☐	...	☐	...	☐	...	☐

* Considered to be uninformative for the purposes of our analysis

Box 3. GP-patient communication items used in the patient experience survey

Ratings by trained external raters

From the video recorded consultations for which the patient had completed the communication items on the questionnaire, we sampled 56 consultations for rating by experienced trained clinical raters. Raters scored each of the selected consultations using the Global Consultation Rating Scale (GCRS). [121] We designed the GCRS to assess the effectiveness of communication across an entire GP–patient consultation; it is based on the widely used Calgary-Cambridge guide to the medical interview. [122, 123] The instrument provides a basis for raters to score each consultation in twelve domains (including gathering information, building the relationship, providing structure and achieving a shared understanding), and results in a final score between 0 and 10 (See Appendix 12). Raters were GPs experienced in the teaching of communication skills; all attended a two-hour training session on GCRS delivered by one of the original authors of the Calgary-Cambridge guide (Jonathan Silverman). We used four raters for each consultation to increase reliability. Raters accessed videos via a secure online portal. Each rater scored consultations in a different random order to minimise the consequences of any order effects, and the same raters were used for all consultations. A simple mean of the four raters was calculated for each consultation and used in subsequent analyses.

From the rating of 56 consultations we expected 80% power (0.05 significance level) to detect a correlation coefficient of 0.37. To obtain the strongest correlation we designed our sampling strategy to include consultations with a wide range of scores: 28 (half) from those where all patient responses to the seven communication items were either good or very good, and 28 (half) where at least one rating was less than good. For the 28 ‘less than good’ consultations, we selected those with the lowest patient communication scores. The 28 ‘good’ consultations were selected at random. We placed a restriction on the selection of consultations which barred the inclusion of more than two consultations involving the same GP.

Statistical analyses

Reliability of GCRS scores

We assessed the reliability of the GCRS scores by fitting a mixed effect linear regression model to the 224 individual ratings (four ratings of 56 consultations). We anticipated that some raters would give systematically higher scores than others, resulting in an inflation of the within consultation variance. As the same four raters were used to rate all 56 consultations this source of variation did not contribute to the reliability, as it manifests itself as a fixed offset in the mean consultation rating used in the analysis outlined below. Thus a categorical fixed effect was included for rater in the models to account for this source of variance. The model additionally had a random intercept for consultation. In this model the variance of the random intercept represents the between consultation variance (σ_b^2) and the residual variance represents the within consultation, between rater, variance (σ_w^2) in ratings (after accounting for systematic differences between raters). The reliability (λ_{GCRS}) of the mean GCRS rating is then given by

$$\lambda_{GCRS} = \frac{\sigma_b^2}{\sigma_b^2 + \frac{\sigma_w^2}{4}} \tag{1}$$

Consultation scores

The 56 consultations selected for rating were used to explore the association between GP's ratings of their own doctor-patient communication, patient ratings of communication, and the scores given by trained raters. The much larger sample of all videoed consultations was used to explore the association between GP and patient scores.

The association between patient scores and trained clinical raters' scores

We explored the association between patient ratings and the ratings obtained by trained raters using a simple correlation coefficient and scatter plot. This coefficient can be corrected to account for the attenuation produced by the less than perfect reliability of the GCRS rating by multiplying by $\sqrt{\lambda_{GCRS}}$. Consideration was given to adjusting for patient socio-demographic characteristics only if this resulted in reduced standard errors; however, this was not the case and so unadjusted results are shown. Because of potential concerns over normality assumptions, bootstrapping was used with 1000 bootstrap samples. To account for the non-independence of observations due to some GPs being represented twice, we performed the bootstrap sampling clustered by GP. Finally, we illustrated the relationship between single consultation ratings and GP ratings made up of many individual patient ratings by simulating scores for 100 hypothetical GPs with a range of communication skills as measured by GCRS. The patient ratings for a given GCRS score were drawn from an appropriate distribution, informed by the findings of the observational work, and then, for each GP, mean patient scores were calculated for 1, 10, 30 and 100 patients.

The association between GP scores and trained clinical raters' scores

We explored the association between GP ratings of their performance and the ratings obtained by trained raters by calculating correlation coefficients. Consideration was given to adjusting for patient socio-demographic characteristics only if this resulted in reduced standard errors; however, this was not the case and so unadjusted results are shown. Because of potential concerns over normality assumptions, bootstrapping was used, in this case with 500 bootstrap samples. Again, to account for the non-independence of observations due to some doctors being represented twice, we performed the bootstrap sampling clustered by doctor.

The association between GP scores and patient scores

To compare GP and patient scores we used all available consultations. Firstly, we carried out a correlation analysis, as above. Subsequently, we conducted a regression analysis with doctor rating as the outcome, adjusting for patient age, gender, ethnicity and self-rated health. Finally, to evaluate the within-doctor association between patient and doctor scores, we augmented the previous model with a random effect for doctor. This final model accounted for the fact that some doctors may, in general, be more generous or more critical that other doctors. Standardised regression coefficients (betas) are reported being directly comparable to (and in the case of models with a single exposure equal to) correlation coefficients. As above, clustered bootstrapping was used for all analysis.

All analysis was carried out using Stata v.13.1 (StataCorp, Texas, USA).

Ethical approval for the study was obtained from the NRES Committee East of England – Hertfordshire on 11 October 2011 (ref: 11/EE/0353)

Results

Consultations were videoed with 45 participating GPs from 13 general practices. During the period of study a total of 908 patients had face-to-face consultations with participating doctors. Of these, 167 (18.4%) were ineligible (mostly children), and 529 completed a questionnaire (71.4% response rate) (see Figure 3 for details of sampling for these analyses). A further 27 (5.1%) consultations were excluded from our analyses due to missing data. The videos selected for rating using GCRS came from all 13 general practices and represented 37 GPs. One further consultation was excluded from our analysis of how GP and rater scores compared, due to a rated consultation missing the communication score from the GP.

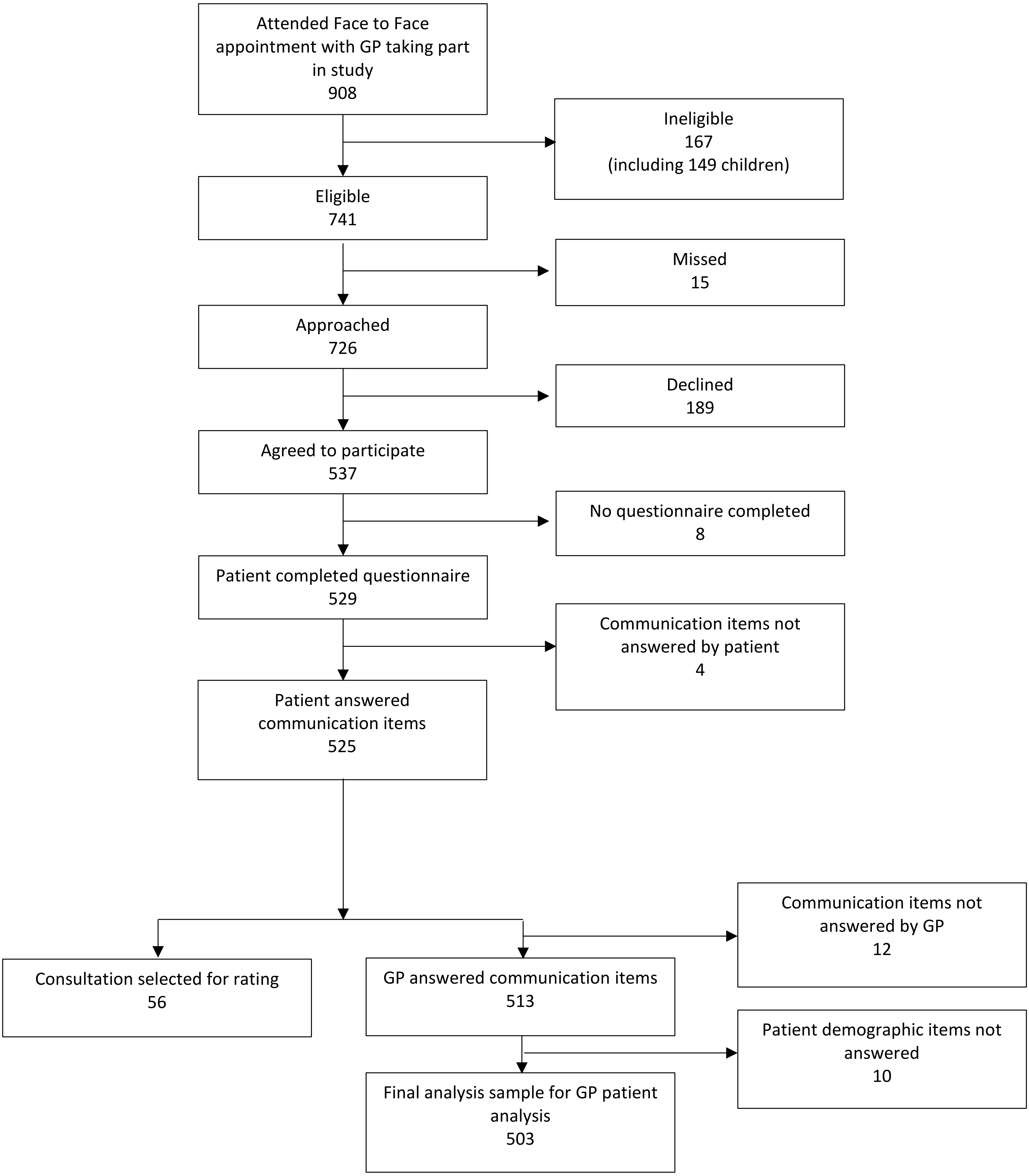


Figure 3. Flow chart illustrating the recruitment and participation of patients

Table 2 shows the self-reported demographics of those patients who completed a questionnaire, along with those whose consultation was selected for rating by trained raters. Men, 18-24 year olds and Asian patients were somewhat more likely to have been selected to have their consultations rated.

Table 2. Self-reported demographics for patients who completed a questionnaire and those selected for consultations to be rated by trained raters.

		Completed questionnaire		Rated consultations	
		n	%	n	%
Sex	Male	212	40.15	26	46.43
	Female	316	59.85	30	53.57
Age	18-24	39	7.41	10	18.18
	25-34	78	14.83	7	12.73
	35-44	64	12.17	7	12.73
	45-54	82	15.59	4	7.27
	55-64	85	16.16	8	14.55
	65-74	103	19.58	7	12.73
	75-84	60	11.41	8	14.55
	85+	15	2.85	4	7.27
Self-rated health	Excellent	50	9.51	3	5.36
	Very good	173	32.89	14	25
	Good	182	34.60	23	41.07
	Fair	83	15.78	13	23.21
	Poor	38	7.22	3	5.36
Ethnicity	White	474	90.98	44	81.48
	Mixed	5	0.96	1	1.85
	Asian or Asian British	15	2.88	6	11.11
	Black or Black British	22	4.22	1	1.85
	Chinese	4	0.77	1	1.85
	Other	1	0.19	1	1.85

Reliability of GCRS scores

The distribution of patient scores and GCRS ratings is shown in Figure 4. Patient scores were highly skewed: the most common score was 100 out of a possible 100 (i.e. very good for all reported communication items: found for 21/56 consultations). The median score was 91 (IQR 71-100), and the lowest score reported 31/100. In contrast, the GCRS ratings are reasonably symmetrical: the median GPPS score was 4.3 (IQR 3.6-5.5), and scores ranged from 2.2 to 6.8 out of a possible 10. From the mixed modelling of GCRS ratings (adjusted for rater) the estimated variances were 1.01 between consultations, and 1.18 within consultations. Reliability for the mean of four ratings was 0.77.

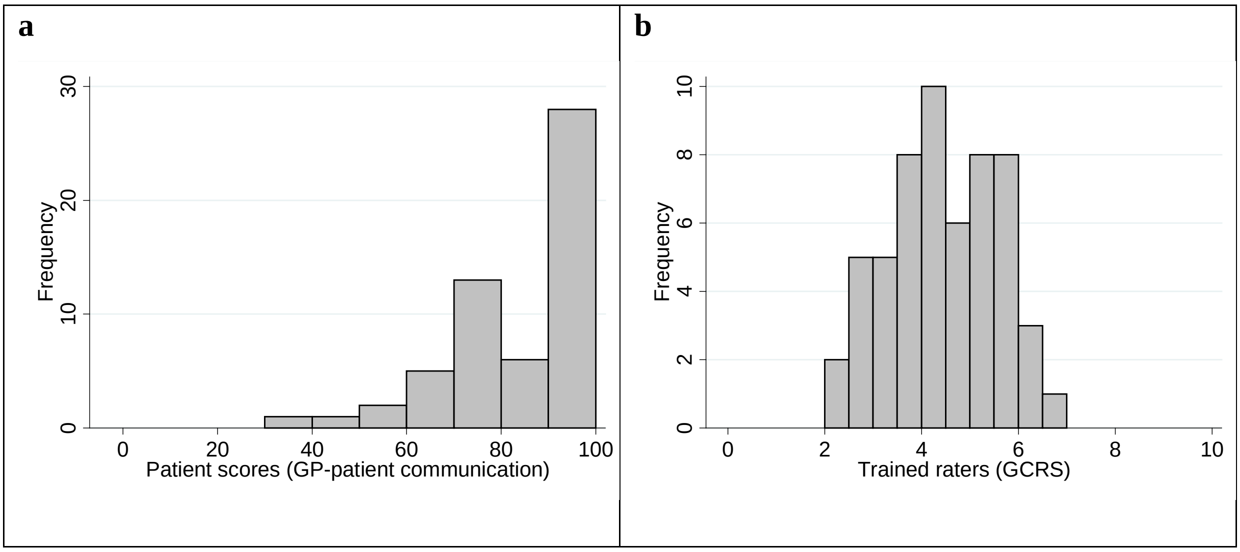


Figure 4. Distribution of patient scores based on GP patient survey items and ratings by trained raters on the GCRS

The association between patient scores and trained clinical raters' scores

Figure 5 shows patient scores plotted against average GCRS ratings for each consultation. There is weak evidence (p=0.054) of an association between patient scores and GCRS ratings, with a correlation coefficient of 0.29. This increases to 0.33 when corrected for attenuation due to the imperfect reliability of the mean GCRS rating. When trained raters assessed communication within a consultation to be of a high standard, patients tended to do the same (with the exception of a single outlying low patient score). However, when trained raters judged communication within a consultation to be of a poor standard, patients reported communication as anything from poor to very good.

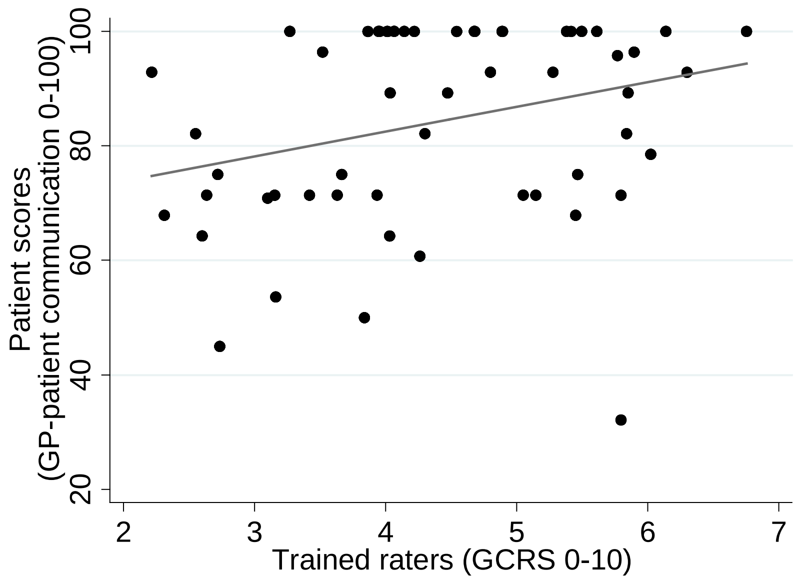


Figure 5. Scatterplot comparing patient scores based on GP patient survey items and ratings by trained raters on the GCRS

Figure 6 shows the results of our simulation study, which is based on a hypothetical set of consultations with a range of trained rater scores (GCRS). For each GCRS score we defined a range of possible simulated patient scores, shown by the shaded grey area in Figure 6. The lower limit of these simulated patient scores increased as GCRS score increased. However, the upper limit of simulated patient scores was set at 100 for all possible GCRS scores in the simulation. For any given GCRS score we allowed patient scores to take any value in this range, with equal probability. The simulation is designed for illustrative purposes only and is not intended to accurately reflect the findings presented here. Panel A, designed to be reminiscent of Figure 5, shows what would be observed with just a single patient score per GP, i.e. a weak correlation between patient rating and communication skill. The remaining panels illustrate the effect of combining scores (taking the mean) from multiple consultations, rather than using a single rating. As the number of patient ratings taken increases, the correlation between trained rater scores and patient scores gets stronger. When the number of consultations are 30 this correlation becomes very strong (rho=0.97), becoming stronger still when n=100.

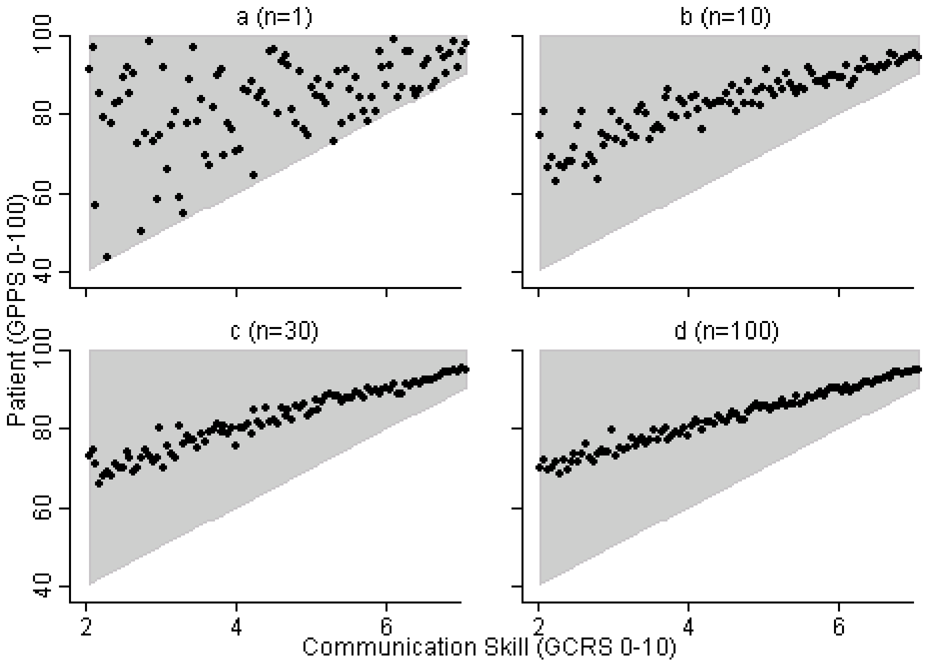


Figure 6. Simulated GP communication scores based on different numbers of patient ratings. The grey areas show the possible individual patient scores that could be given for any particular level of communication competence, as assessed by GCRS

The association between GP scores and trained clinical raters' scores

Histograms showing the distribution of scores given to the 55 consultations with both GP scores and trained rater scores are shown in Figure 7a and b. Both distributions are reasonably symmetrical. The GCRS scores cover a wide range of the possible values; in contrast, the GP scores of their own performance were all over 50 out of 100, indicating that no GP scored themselves poor or very poor consistently across the domains for any one consultation. Scatterplots comparing GP scores to the GCRS are shown in Figure 8a. The wide scatter is reflected in the low correlation coefficients shown in Table 3 with no evidence that GP scores are associated with the scores of trained raters using the GCRS (p=0.721). Because only a small number of consultations were rated, confidence intervals are wide. However, it is of particular note that the upper confidence interval is below, 0.25 indicating that moderate or strong correlations between GP scores and rater scores are highly unlikely to be consistent with these data.

Table 3. Correlation coefficients/standardised regression coefficients used to examine the association between physician scores and the scores given by the two sets of trained raters, and by patients

	Trained raters GCRS (n=55)		Patients (n=503)	
	Correlation coefficient (95% CI)	P-value	Correlation coefficient/ Standardise regression coefficients (95% CI)	P-value
Global association	-0.052 (-0.336, 0.232)	0.721	0.009 (-0.086, 0.104)	0.854
Within GP association	N/A		0.025 (-0.060, 0.110)	0.565
Within GP association adjusted for patient socio-demographics	N/A		0.023 (-0.064, 0.110)	0.608

The association between GP scores and patient scores

Figure 7c and 7d show the distribution of GP scores and patient scores for all consultations where both are present (along with patient socio-demographic information). The distribution of GP scores is similar to that seen in the selection used for rating. In contrast, the distribution of patient scores is highly skewed with 63.4% of patients giving the maximum score of 100. A scatter plot comparing GP scores with patient scores of the same consultation in shown in Figure 8b. The skewed nature of patient scores is evident in this Figure, which also shows that whilst GPs do not often score themselves lower than 50, they on average give themselves lower scores than patients. The lack of any clear relationship in this Figure is reflected in the very low correlation coefficient shown in Table 3, again with no evidence of an association (p=0.854). The lack of association persists when considering within-GP associations and when further adjusting for patient demographics. Due to the increased sample size confidence intervals are tighter than those found when comparing with rater scores such that only very weak correlations between GP and patient scores would be consistent with these data.

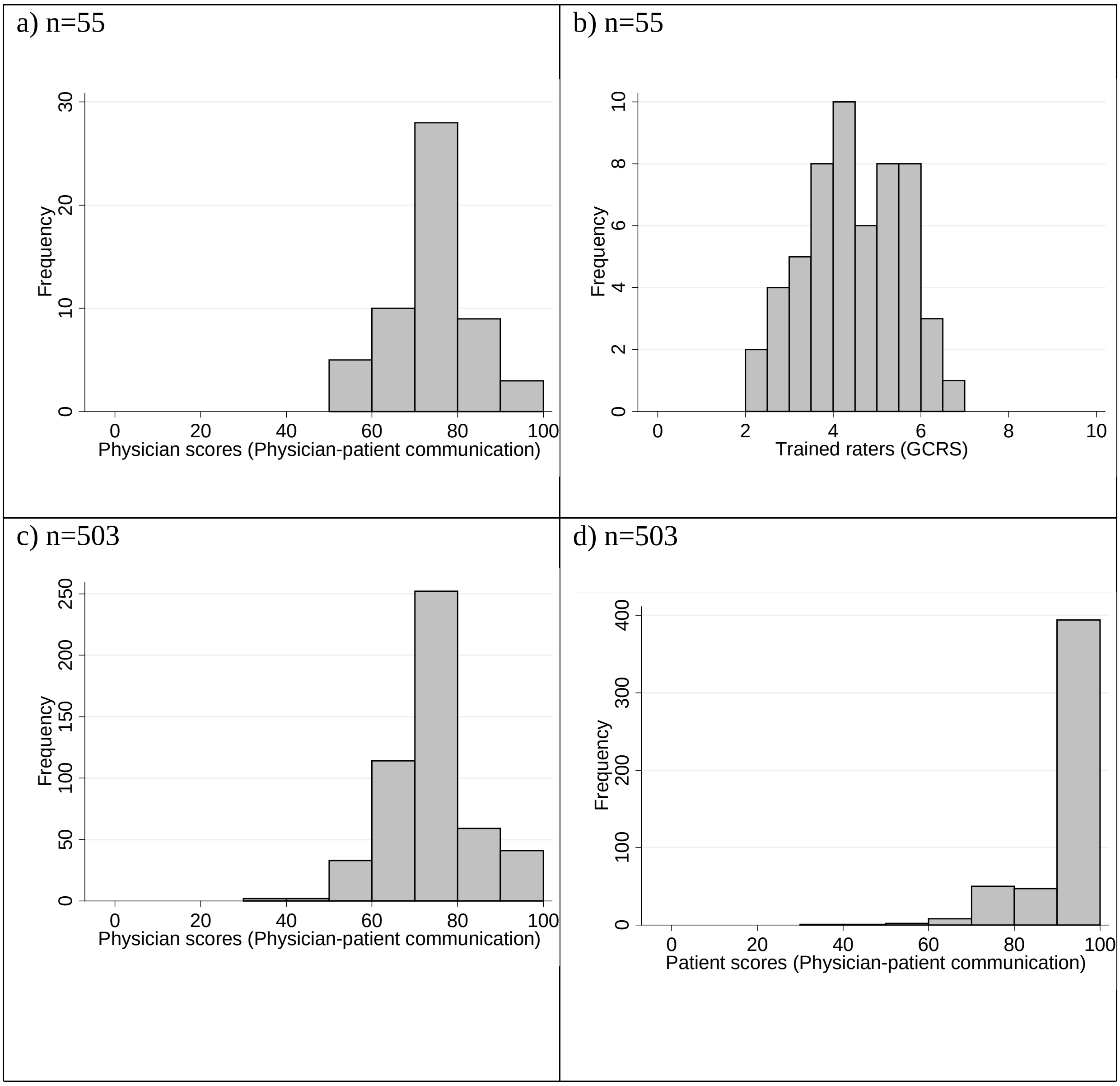
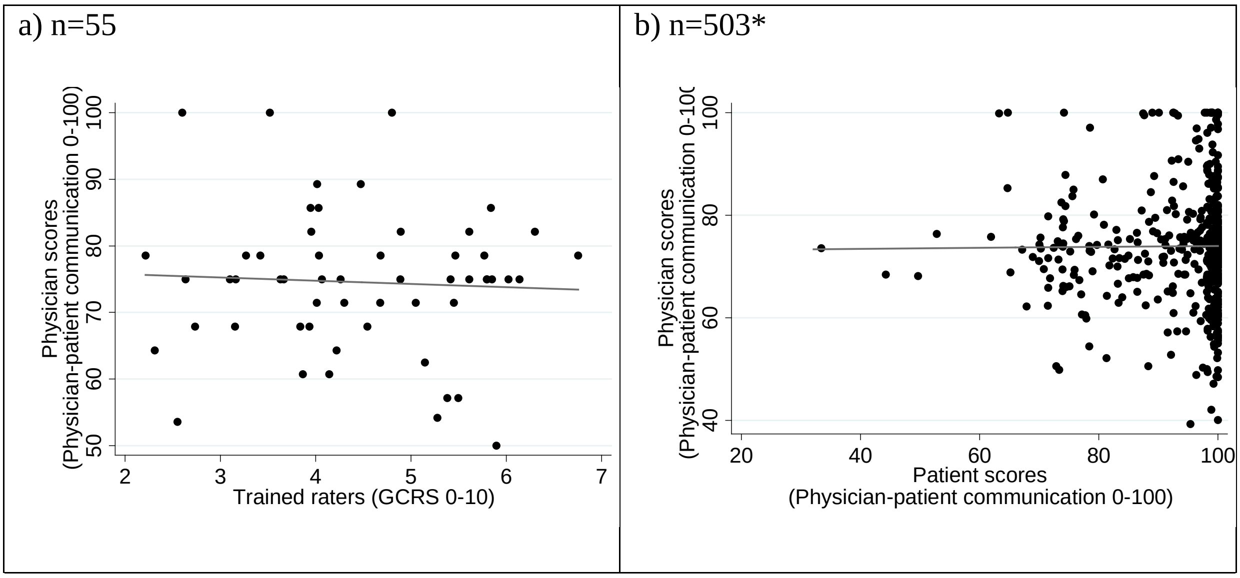


Figure 7. Distribution of scores given to consultations by GPs scoring themselves (a and c), raters using the GCRS scale (b), and patients (d). Panels a and b apply only to those consultations rated by trained raters, whilst panels c and d relates to all consult



Note: In each case the grey line is a line of best fit. Panel (a) applies only to those consultations rated by trained raters, whilst panel (b) relates to all consultations where physician and patient scores were recorded along with patient demographics.

Figure 8. Scatterplots illustrating the association between GP scores and a) trained rater scores using the GCRS scale, and b) patients

Discussion

We found a modest correlation between patients' and trained raters' assessments of the quality of communication in GP-patient consultations. This suggests that there is an association between patient ratings of communication and professionally-defined standards of care. Importantly, when trained raters identified communication as good, patients tended to agree with this. However, when trained raters identified communication as poor, patients ranged in their assessments of communication from poor to very good. By contrast, we found no evidence of an association between GPs' and trained raters' assessments of communication performance, and no evidence of an association between GPs' and patients' assessments of communication quality.

The first aim of this workstream was to explore how far patient ratings reflect accepted professional standards of communication. Our findings suggest that, whilst trained raters' and patient's tend to agree what good communication looks like in a consultation, clinical raters are more likely than patients to judge communication as poor. We outline two possible mechanisms driving this divergence: both assume that raters' assessments of communication quality are the 'gold standard' (an issue we discuss further below). The first mechanism arises from the well-known phenomenon of skewed patients' ratings, with a large proportion of patients rating communication as "very good". [119, 124, 125] By contrast, GCRS ratings tend to cluster around the middle scores available to raters. It is therefore possible that the presence of ceiling effects inherent in the patient question items may artificially constrain the responses patients would like to give, preventing them from being able to distinguish the very best consultations from those they judge to be simply good. If our observed pattern is attributable to ceiling effects, this implies a weak correlation between underlying 'true' patient opinion (not the reported opinion expressed via available survey instruments) and trained rater scores. Thus, this mechanism requires that patients differ from raters in their views of what good or poor communication in a consultation looks like. As a result, the more positive patient opinion is "held back" by only being able to choose questionnaire options ranging from very poor to very good (and not, for example, 'excellent'), despite extensive instrument development. [126]

However, the second – and we argue more plausible – mechanism is that there are wider factors at play which inhibit some patients from assigning poor scores to consultations which they do perceive as involving poor communication. It is important to note that any such inhibition would have to apply unevenly between patients to explain the range of patient scores seen for consultations rated as poor by the trained raters: whilst some patients *are* able to choose 'poor' as an option, others feel less able to do so. For this mechanism to be driving our observed pattern, patients' 'true' opinion would be more strongly correlated with trained raters' opinion, with general agreement between patients and raters about what good or poor communication looks like. Such a phenomenon may lead to an overestimate of doctors' "true" patient experience scores, whereas the former mechanism (in which patients are constrained in their choice of responses) would lead to an underestimate of doctor performance.

Whilst we are unable to determine the relative contribution of either mechanism from the methodology of this current workstream, there is existing evidence that patients may be inhibited in their judgements of care. In particular, qualitative research has identified a number of psychological and social factors that suggest patients struggle to criticise GPs' performance in surveys. For example, an investigation into how patients evaluated community mental health services found that they frequently avoided giving negative scores on experience questionnaires: instead, allowances for poor care were constructed by referencing their perceptions of the duties and culpabilities of health care providers. [90] Similarly, patients undergoing elective orthopaedic surgery re-interpreted their experiences in a positive light as a result of feelings of dependency on their health care providers, and a perceived need to maintain constructive relationships with GPs. [89] A tendency to excuse rather than report poor care has also been identified in breast cancer patients. [127] These findings were confirmed in the qualitative research we undertook with our sample, as previously discussed in Chapter 2.

The lack of association between GPs own ratings of their performance and both trained raters' and patients' assessments of the same echoes previous research, which has identified gaps in the perceptions of doctors' and others' evaluations. [114-116] The absence of agreement between GPs and trained raters, and between GPs and patients, suggests that their assessments of what constitutes a "good" consultation may vary. This has potentially important implications for the utility of GP self-reflection in developing their clinical practice. Reflective practice has become a core part of continuing professional development over the years, and the identification of learning needs forms an important aspect of this. [128] Indeed, the collection and consideration of patient feedback is central component of the supporting information required for current medical appraisal and revalidation, and appraisers are required to explore "what [doctors] think the supporting information says about [their] practice and how [they] intend to develop or modify your practice as a result of that reflection." [129] In this study, trained clinical raters (all GPs) were consistently more negative about communication performance than participating GPs were. In seeking to improve communication skills and patient experience, reliance on a doctors' own assessment may not be a robust approach, and there may be an important role for external assessment of communication performance.

Strengths and limitations

It was not necessary to seek a representative sample of practices or GPs for the purposes of this study. Instead, we intentionally approached a sample of practices previously found to be receiving lower patient experience scores for communication. Our sampling strategy was informed by the need to locate consultations patients identified as less than good; the proportion of such consultations is small, so to increase study efficiency we deliberately approached practices who had received lower scores for communication in the national GP

Patient Survey. Not all GPs in every practice took part, and it is possible that the GPs who did so were more confident in their ability to communicate with patients.

Our patient consent rate was 71.4% of eligible patients. The research team missed only a small number of patients (2.0% (15/741) of those eligible), so exclusions predominantly reflect those who did not consent to participate. Recorded consultations concerning some medical conditions may be under-represented as participants may have been more likely to decline being video recorded. However, participants' age, gender, self-rated health and ethnicity were broadly representative of the population attending general practices.
We assessed communication using two well validated instruments: the GP Patient Survey items for patients and GPs, and the GCRS for trained raters. [121, 126] The GCRS was derived from the Calgary-Cambridge guide, which is used widely for communication skills training, and represents agreed professional norms of high quality communication. [122, 123, 130] Recently, the question has arisen as to how and whether trained raters take account of contextual factors in assessing the communication skills of GPs, for example, by allowing variations from 'accepted practice' when scoring performance in particular situations. [131, 132] However, the GCRS has been explicitly designed to focus only on the consultation process, and contains no task-based items which may be context-specific (such as requiring a rating for specific physical examinations). Additionally, it enables raters to choose 'not applicable' where necessary: in fact, this was rarely endorsed by raters in this study.

As mentioned above, in drawing conclusions about the meaning of patients' and GPs' ratings of communication quality, we have positioned the trained raters as being "the gold standard". This is not to suggest raters are more valued or competent assessors of communication than patients or GPs, but simply to use them as representative of professionally-agreed norms of behaviour against which to judge patient and participating doctor evaluations of communication. In doing so, we are able to provide evidence that patient assessments tap in to the same underlying construct of communication drawn upon by trained raters, but also that patients are less likely to judge consultations as poor. We are also able to provide evidence that GP assessments of their own performance do not appear to be associated with the same ideas of what "good" communication looks like for trained raters.

Conclusions

Our findings support observations that patients may be inhibited in criticising doctors' performances. If indeed patients are reluctant to give lower ratings which truly reflect their experience, mean survey scores may be overestimates of performance. We therefore suggest that the practice of taking mean survey scores at face value and assuming that they provide a realistic reflection of absolute performance level of either GPs or their practices is inadvisable, as such scores are likely to be biased. However, the use of relative rankings to identify GPs who are better or poorer at communicating with patients may be an acceptable approach to benchmarking performance, as long as statistically reliable figures are obtained. Previous research has demonstrated that the GP Patient Survey communication questions can differentiate between the performance of GPs and practices, as long as an adequate sample size is used to achieve acceptable statistical reliability. [118, 133] This was confirmed by our simulation: with sufficient patient scores a strong correlation between patient rating and competency will be observed. In the use of patient experience scores as quality indicators, our findings suggest that it is therefore possible to (a) trust aggregated patients scores that meet traditional standards of reliability as valid measures of comparative performance with respect to communication and (b) trust relatively low mean patient ratings. However, crucially, we cannot necessarily assume that a high mean patient rating means all is well.

GP assessments of their own standards of communication were poorly associated both with professionally-defined norms of communication and with patients' own assessments of what happened in a consultation. Taken together with our findings on the importance of rater feedback on identifying consultations where communication is less than ideal, our findings suggest that there may be a current gap in the use of external assessments of communication competence.

Chapter 4. Ability of patients to see the clinician of their choice

Abstract

Background

This chapter describes analyses of data from the GP Patient Survey investigating which patients have a preference for seeing a particular GP, and how successful they are in seeing that doctor. We report these trends over a four year period. In addition, we undertook analyses to examine whether patients' expectations of whom they wished to see and who they did see (a doctor or a nurse) influenced their assessment of the consultation.

Methods

Analysis of data from the GP Patient Survey

Results

The majority of patients have a particular GP whom they prefer to see. This increases from just over 50% in 18-25 year olds to 80% in patients over 75. Of those patients who have a preference to see a particular GP, 30% were not able to see that doctor easily in 2010/11. That percentage has been rising year on year, to 39% in 2013/14, indicating substantial problems in patients getting to see the doctor of their choice. Patients who saw a nurse when they wanted to see a GP gave scores for communication with the nurse that were substantially lower (adjusted difference 5.99%, 95% CI 5.71-6.28) than patients who wanted to see a nurse in the first place.

Conclusions

Patients' ability to see a doctor of their choice is seriously compromised, with a high proportion of patients who have a preference for a particular doctor unable to see that doctor on a regular basis. This is a significant quality problem for the NHS.

Introduction and rationale for the study

This chapter relates specifically to the fourth aim of the programme: "To understand better patients' responses to questions on communication and seeing a doctor of their choice". In all the rest of the programme, we have focused on communication between doctors and patients. In this chapter we present our analyses of questions in the GP Patient Survey which relate to patients' ability to see the doctor of their choice.

Continuity of care, specifically relational continuity, is valued by patients and is a core value of general practice. Nevertheless, changes to practice organisation and staffing (including targets relating to improved access) have all combined to make it more difficult for patients to see a regular doctor. There are no routinely collected measures of continuity of care. However, there are two questions which have remained largely unchanged in the GP Patient Survey for several years which are relevant to continuity of care. These are shown in Table 4.

Table 4. Questions on choice of doctor as they appear in the GP Patient Survey

Q15. Is there a particular doctor you prefer to see at your GP surgery or health centre?
☐ Yes ….. Please go to Q16
☐ No …… Please go to Section F
☐ There is usually one doctor in my GP surgery or health centre ….. Please go to section F
Q16. How often do you see the doctor you prefer to see?
☐ Always or almost always
☐ A lot of the time
☐ Some of the time
☐ Never or almost never
☐ Not tried at this GP surgery or health centre

Note: Prior to July 2011. From July 2011 onwards the first question was changed to "Is there a particular GP you usually prefer to see or speak to?" and the second question was changed to "How often do you see or speak to the GP you prefer?". Response options remained the same throughout apart from the removal of the words "or health centre" from the final response option for each question.

The wording of this question recognises that not all patients want to see a particular doctor, and it enables the ability to see a particular doctor to be assessed among patients who have that preference. These questions are not a direct measure of continuity of care, but combine elements of continuity with an element of patient choice.

The aim of the analysis of these questions was, first, to identify which patients most valued having a particular doctor and, secondly, to examine the extent to which patients were able to see the doctor of their choice. Because these questions have remained stable for some years, we were also able to examine trends over time.

We also include in this chapter some analyses of patient experience with practice nurses. While this was not part of our original programme of work, they are of interest in their own right, and also give us the opportunity to examine patient responses when they wished to see a doctor but the appointment they were given was with a nurse and the corollary, when they wished to see a nurse but were given an appointment with a doctor.

Methods

Three sets of analyses are presented in this chapter:

1. Determination of which patients express a preference for and manage to see a doctor of their choice.
2. Examination of trends in the proportion of patients able to see a doctor of their choice.
3. The association between patient rating of communication and the mismatch between the type of appointment wanted and the type of appointment received.

Analysis 1

Data from the 2009/10 GP Patient Survey were used for the analyses of which patients had a preference for, and succeeded in seeing, a particular doctor. The results presented are a summary of those published in the British Journal of General Practice. [134]

Responses to how often patients were able to see their preferred doctor (where they expressed a preference for doing so) were dichotomised into a Yes (“always or almost always” or “a lot of the time”) / No (“some of the time” or “never or almost never”) measure. Survey weights were developed by Ipsos MORI (the survey provider) and were used in our analysis to account for the complex survey design and non-response in prevalence estimates of preference for and success in seeing a preferred doctor. These weights employed rim weighting with 2 rims: a) age-by- gender (8*2 levels) and b) practice (8,362 levels). Crude and multivariate logistic regression models were used to examine the association between various patient and practice characteristics with preference for and success in seeing a preferred doctor, separately. We adjusted for gender, age group, ethnicity, deprivation quintile, self-reported chronic medical or psychological/emotional condition, number of practice doctors, and the type of appointments requested by the patient in the previous six months. Crude models made use of the weights and adjusted standard errors to account for the survey design. Multivariate models did not make use of the weights but did include random intercepts for practice to account for clustering of patients within practices and to better distinguish the experiences and preferences of patient subgroups from general variation in continuity at practice level.

Although patient registration with a given practice is largely determined by geographical proximity, some patients might choose to register with a smaller practice specifically to get better continuity of care, in which case it would not have been appropriate to adjust socio-demographic associations for practice size. For this reason we performed a sensitivity analysis excluding the number of practice doctors. The results were very similar, for which reason data are not shown. Stata v.11 (StataCorp, Texas, USA) was used for the descriptive analyses and SAS v.9.2 for the regression analyses.

Analysis 2

In this analysis we used data from four years of the GP Patient Survey (2010/11 to 2013/14), where the questions addressing the ability to see a preferred GP have remained unchanged. We present annual national figures after applying the survey design and non-response weights such that percentages are representative of the national population rather than respondents to the survey.

Analysis 3

For this analysis we used data from the 2013/14 GP Patient Survey. This included more detailed questions on appointments than previous surveys and, in particular, we were able to analyse the responses of patients seeing a nurse on their last visit, comparing those who contacted the practice wanting to see a nurse from those whose original request had been to see a doctor. Similarly, we were able to analyse the responses of patients seeing a doctor on their last visit, comparing those who wanted to see a doctor from those whose original request had been to see a nurse.

We first present a descriptive analysis to examine the extent to which the type of appointment people obtained was the same as of different to the type of appointment they wanted. This was done by comparing responses to the question “Last time you wanted to see or speak to a GP or nurse from your GP surgery: What did you want to do?” with responses to the question “What type of appointment did you get? I got an appointment…”. For both questions response options allowed patients to indicate that they wanted to/got to; see a GP at the surgery, see a nurse at the surgery, speak to a GP on the phone, speak to a nurse on the phone, have a home visit. Additionally, when asking about what they wanted, there was an option to state “I didn’t mind /wasn’t sure what I wanted”. Because patients often endorsed more than one response we reduced responses to the first question into five categories:

1. Those who either wanted to see or speak to a GP (or both)
2. Those who either wanted to see or speak to a nurse (or both)
3. Those who wanted a home visit
4. Those that weren’t sure or didn’t mind
5. Those who wanted more than one of the previous four categories.

For the second question the same categories were used, excluding the “Didn't mind/wasn't sure” category as this was not an option for that question. We then cross tabulated what people wanted with what they got, again using the design and non-response weights.

Finally, we considered whether reported nurse-patient communication and doctor-patient communication varied according to any mismatch between what people wanted and what they got in terms of who the appointment was with and what type of appointment it was. Two separate analyses were performed (one for nurse communication and one for doctor communication). Each analysis was restricted to those reporting that they had had an appointment with the appropriate clinician on their last visit to their GP surgery. For relative simplicity this analysis was further restricted to those who only endorsed one box for both the questions on what type of appointment they wanted and the question of what they got. Eight categories were created (for each analysis) covering the various combinations of seeing or speaking to someone and whether the person they wanted to see or speak to was a GP. Composite nurse-patient communication and doctor-patient communication scores, between

0 and 100, were created in an identical way to those described earlier in this report, except using responses to the equivalent question about nurses where appropriate. Crude differences between the categories were estimated using linear regression (restricted to those who had complete information for age, gender, ethnicity, confidence in managing their own health, the presence of a long standing heath condition, and deprivation). Mixed effects linear regression was then used for an adjusted analysis including age, gender, ethnicity, confidence in managing their own health, the presence of a long standing heath condition, and deprivation. Practice was included as a random effect (intercept).

Results

The overall response rate to the 2009/10 GP Patient Survey was 39% with 2,169,718 completed responses from patients in 8,362 practices.

Analysis 1a. Preference for seeing a particular doctor

Two per cent of patients reported that there was only one GP in their practice. After excluding those patients from further analysis, 62% of patients reported having a preference for seeing a particular doctor (Table 5). Such a preference varied across patient groups (Table 5) and was higher in women (68% vs. 56% in men), older patients (52% for age group 18-24 increasing to 81% for age group 75-84), those with chronic medical or psychological/ emotional conditions (75% and 78% respectively) compared to those without (52% and 61% respectively) and those living in more affluent areas (from 60% to 64% for most to least deprived patients). Preference for seeing a particular doctor ranged from 47% to 65% of respondents across the 16 ethnic groups, and increased with the number of practice GPs (58% for practices with 2 GPs increasing to 63% for practices with 6-9 GPs). Preference for seeing a particular doctor was higher in patients who had recently requested only non-urgent appointments in the previous six months (68%) compared to patients who had requested only urgent appointments (58%). The crude odds ratios (Table 5) reflect the associations described above. All associations are stronger than would be expected by chance ($p<0.001$).

In multivariate analysis there was strong evidence that differences exist in the preference for seeing a particular doctor across all socio-demographic groups after adjusting for other factors ($p<0.001$ for all variables) (Table 5). This preference was commoner among women (OR=1.50), older people (OR=1.71 for age group 74-85 compared to age group 55-64), respondents suffering from a chronic medical (OR=1.87) or psychological/emotional condition (OR=1.59), and those from more affluent areas (OR=0.84 for most deprived compared to most affluent areas). Patients from South-Asian ethnic groups (Bangladeshi, Indian, Pakistani and ‘Any other Asian’) had substantially higher preference for seeing a particular doctor (OR=1.74, 1.49, 1.49, and 1.28 respectively compared to White British). Patients were more likely to express such preference if they were registered with practices with greater number of GPs (OR=1.3 for patients registered with practices with 6-9 GPs compared to patients registered with practices with 2 GPs) and if they had sought non-urgent appointments (OR=1.4 patients seeking non-urgent appointments only compared to patients seeking urgent appointments only).

Table 5. Prevalence and odds ratios for having preference for seeing a particular doctor

		Weighted prevalence* **(95% CI)**	**Crude Weighted ORs*** **(95% CI)**	**Adjusted ORs**** **(95% CI)**
All survey respondents		62.2 (61.9,62.4)	N/A	N/A
Gender				
Male		56.3 (56.1,56.6)	Ref	Ref
Female		67.5 (67.2,67.7)	1.60 (1.59,1.61)	1.50 (1.49,1.52)
Age group				
18-24		51.7 (51.1,52.2)	0.49 (0.48,0.50)	0.65 (0.64,0.66)
25-34		51.0 (50.6,51.3)	0.48 (0.47,0.48)	0.55 (0.54,0.56)
35-44		56.0 (55.6,56.3)	0.58 (0.58,0.59)	0.66 (0.65,0.67)
45-54		61.4 (61.1,61.7)	0.73 (0.72,0.74)	0.79 (0.78,0.80)
55-64		68.6 (68.3,68.9)	Ref	Ref
65-74		76.0 (75.8,76.3)	1.45 (1.44,1.47)	1.36 (1.35,1.38)
75-84		81.1 (80.8,81.4)	1.97 (1.94,2.00)	1.71 (1.69,1.74)
Over 85		80.0 (79.5,80.4)	1.83 (1.78,1.88)	1.54 (1.50,1.58)
Ethnic group (ONS 6)	**Ethnic group (ONS16)**			
White	White British	62.9 (62.6,63.2)	Ref	Ref
	Irish	65.1 (64.2,65.9)	1.10 (1.06,1.14)	0.97 (0.94,1.00)
	Any other white	57.5 (56.9,58.1)	0.80 (0.78,0.82)	1.03 (1.01,1.05)
Mixed	White and Black Caribbean	56.8 (54.8,58.9)	0.78 (0.72,0.84)	1.05 (0.97,1.14)
	White and Black African	52.2 (49.6,54.7)	0.64 (0.58,0.71)	0.92 (0.84,1.02)
	White and Asian	56.7 (53.8,59.6)	0.77 (0.69,0.87)	1.07 (0.99,1.16)
	Any other Mixed	59.7 (57.7,61.7)	0.88 (0.81,0.95)	1.09 (1.02,1.18)
South-Asian	Indian	63.0 (62.2,63.9)	1.01 (0.97,1.04)	1.49 (1.45,1.53)
	Pakistani	61.4 (60.4,62.4)	0.94 (0.90,0.98)	1.49 (1.43,1.54)
	Bangladeshi	61.7 (60.2,63.2)	0.95 (0.89,1.01)	1.74 (1.64,1.84)
	Any other Asian	59.0 (57.9,60.1)	0.85 (0.81,0.89)	1.28 (1.23,1.33)
Black	Black Caribbean	61.9 (60.9,62.8)	0.96 (0.92,1.00)	1.14 (1.10,1.18)
	Black African	47.3 (46.4,48.2)	0.53 (0.51,0.55)	0.81 (0.78,0.83)
	Any other Black	59.1 (57.2,61.1)	0.86 (0.79,0.93)	1.08 (0.99,1.17)
Chinese	Chinese	48.5 (47.0,50.0)	0.56 (0.52,0.59)	0.86 (0.81,0.90)
Other	Other ethnic group	58.5 (57.9,59.1)	0.83 (0.81,0.85)	1.14 (1.12,1.17)
Deprivation quintile				
1 (Affluent)		64.1 (63.7,64.5)	Ref	Ref
2		63.3 (62.9,63.6)	0.96 (0.95,0.98)	0.96 (0.95,0.97)
3		62.2 (61.8,62.6)	0.92 (0.90,0.94)	0.92 (0.91,0.93)
4		61.0 (60.6,61.3)	0.87 (0.86,0.89)	0.89 (0.88,0.90)
5 (Deprived)		59.5 (59.1,60.0)	0.82 (0.81,0.84)	0.84 (0.83,0.85)
Presence of self-reported chronic medical condition				
No		52.1 (51.7,52.4)	Ref	Ref
Yes		74.8 (74.5,75.0)	2.73 (2.70,2.76)	1.87 (1.86,1.89)
Presence of self-reported long-standing psychological or emotional condition				
No		61.3 (61.1,61.6)	Ref	Ref
Yes		78.3 (77.9,78.7)	2.28 (2.23,2.33)	1.59 (1.57,1.62)

[continued on next page]

	Weighted prevalence* (95% CI)	Crude Weighted ORs* (95% CI)	Adjusted ORs** (95% CI)
Number of practice GPs			
1GP	56.4 (55.6,57.2)	0.94 (0.90,0.97)	1.01 (0.97,1.05)
2GPs	58.0 (57.4,58.6)	Ref	Ref
3GPs	61.9 (61.3,62.5)	1.18 (1.13,1.22)	1.20 (1.15,1.25)
4GPs	63.5 (62.9,64.1)	1.26 (1.22,1.30)	1.29 (1.24,1.35)
5GPs	63.6 (63.0,64.2)	1.26 (1.22,1.31)	1.31 (1.26,1.37)
6-9GPs	63.0 (62.6,63.4)	1.23 (1.19,1.27)	1.30 (1.25,1.34)
10+ GPS	62.2 (60.7,63.6)	1.19 (1.11,1.27)	1.28 (1.19,1.37)
Type of appointments sought in previous 6 months* **			
No appointment requested	47.5 (47.2,47.8)	0.65 (0.64,0.66)	0.67 (0.66,0.68)
Urgent only	58.3 (58.0,58.6)	Ref	Ref
Non-urgent only	67.5 (67.1,67.9)	1.49 (1.46,1.51)	1.40 (1.39,1.42)
Both urgent and non-urgent	73.5 (73.2,73.7)	1.98 (1.96,2.00)	1.85 (1.83,1.87)

* Estimated from weighted unadjusted analysis. p<0.001 for all association (joint tests for categorical variables).
** Estimated from a single multivariate logistic regression model including all variables which appear in the table plus a random practice intercept. P<0.001 for all associations. (joint tests for categorical variables).
*** Based on patients' response to questions 6 and 9 (see Methods).

Analysis 1b. Ability to see the doctor of the patient's choice

The next analyses are restricted to patients with a preference for seeing a particular doctor. Of these patients, 72% were successful in seeing the doctor they preferred "*always or almost always*" or "*a lot of the time*" – we refer to those two response categories using the term 'most of the time' hereafter (Table 6). The proportion of patients who were successful in seeing their preferred GP most of the time was higher in men (74% vs. 70% in women), older patients (60% for age group 18-24 increasing to 87% for age group 75-84) and those with chronic medical or psychological/emotional conditions (77% and 75% respectively) when compared to those without (66% and 72% respectively). White patients were more likely to be able to see the doctor of their choice compared to most other ethnic groups. More deprived patients were less successful in seeing the doctor they preferred most of the time (67% for most deprived rising to 74% for the least deprived patients). Success in seeing a particular doctor decreased as number of practice GPs increased (79% for practices with 1 GP compared to 69% for practices with 10 or more GPs). Success in seeing a particular doctor was least among patients requesting urgent appointments only (69%), where it was greatest for patients requesting only non-urgent appointments (79%). The crude odds ratios (Table 6) reflect the associations described above. All associations are stronger than would be expected by chance (p<0.001).

In multivariate analysis there was strong evidence that differences in success of seeing a preferred doctor persist after adjusting for other factors ($p<0.001$ for all variables) (Table 6). Women were less likely to be successful than men in seeing the doctor of their preference (OR=0.87). This contrasts with older patients (OR=1.82 for age group 74-85 compared to age group 55-64), those with a chronic medical (OR=1.29) or psychological/emotional condition (OR=1.25), and White patients, all of whom were more likely to be successful compared to their respective reference groups. Success in seeing a preferred doctor was also less common in deprived areas (OR= 0.86 for most deprived compared to most affluent). Patients registered with larger practices were less likely to report that they could see a doctor of their choice most of the time (OR=0.48 for patients registered with practices with 6-9 doctors compared to patients registered with practices with 2 doctors). Patients who requested only non-urgent appointments were more likely to be successful in seeing the doctor they prefer (OR=1.59 compared to patients requesting only urgent appointments).

Table 6. Prevalence and adjusted odds ratios for seeing a preferred doctor most of the time

		Weighted prevalence* (95% CI)	**Crude Weighted ORs* (95% CI)**	**Adjusted ORs** (95% CI)**
All survey respondents		71.8 (71.4,72.1)		
Gender				
Male		73.6 (73.2,74.0)	Ref	Ref
Female		70.2 (69.9,70.6)	0.85 (0.84,0.86)	0.87 (0.86,0.88)
Age group				
18-24		59.8 (59.1,60.5)	0.43 (0.42,0.44)	0.43 (0.42,0.44)
25-34		60.2 (59.7,60.8)	0.44 (0.43,0.44)	0.48 (0.47,0.49)
35-44		63.6 (63.1,64.0)	0.50 (0.49,0.51)	0.54 (0.53,0.55)
45-54		69.8 (69.4,70.3)	0.67 (0.66,0.68)	0.68 (0.67,0.69)
55-64		77.7 (77.3,78.1)	Ref	Ref
65-74		84.3 (84.0,84.7)	1.55 (1.52,1.58)	1.53 (1.50,1.56)
75-84		86.5 (86.1,86.8)	1.84 (1.80,1.87)	1.82 (1.79,1.86)
Over 85		85.3 (84.8,85.7)	1.66 (1.61,1.72)	1.56 (1.51,1.61)
Ethnic group (ONS 6)	**Ethnic group (ONS16)**			
White	White British	73.7 (73.4,74.1)	Ref	Ref
	Irish	74.1 (73.2,75.1)	1.02 (0.97,1.07)	0.90 (0.86,0.94)
	Any other white	66.9 (66.2,67.5)	0.72 (0.70,0.74)	0.85 (0.83,0.88)
Mixed	White and Black Caribbean	61.8 (59.1,64.4)	0.58 (0.52,0.64)	0.90 (0.81,1.00)
	White and Black African	56.6 (53.4,59.8)	0.46 (0.41,0.53)	0.68 (0.60,0.78)
	White and Asian	63.4 (60.7,66.1)	0.62 (0.55,0.69)	0.81 (0.72,0.90)
	Any other Mixed	62.4 (60.1,64.7)	0.59 (0.54,0.65)	0.74 (0.67,0.81)
South-Asian	Indian	60.7 (59.3,62.1)	0.55 (0.52,0.58)	0.73 (0.71,0.76)
	Pakistani	54.4 (52.9,55.9)	0.43 (0.40,0.45)	0.66 (0.63,0.69)
	Bangladeshi	50.2 (48.1,52.3)	0.36 (0.33,0.39)	0.57 (0.53,0.61)
	Any other Asian	56.8 (55.4,58.1)	0.47 (0.44,0.49)	0.59 (0.56,0.62)
Black	Black Caribbean	65.6 (64.2,67.0)	0.68 (0.64,0.72)	0.83 (0.79,0.87)
	Black African	52.3 (50.9,53.8)	0.39 (0.37,0.42)	0.55 (0.53,0.58)
	Any other Black	58.0 (55.3,60.6)	0.49 (0.44,0.55)	0.70 (0.62,0.78)
Chinese	Chinese	56.2 (54.1,58.4)	0.46 (0.42,0.50)	0.55 (0.51,0.59)
Other	Other ethnic group	60.6 (59.9,61.4)	0.55 (0.53,0.57)	0.66 (0.64,0.68)
Deprivation quintile				
1 (Affluent)		74.3 (73.7,74.8)	Ref	Ref
2		74.1 (73.6,74.6)	0.99 (0.97,1.02)	0.99 (0.97,1.01)
3		72.3 (71.8,72.8)	0.90 (0.88,0.93)	0.95 (0.94,0.97)
4		69.6 (69.1,70.1)	0.79 (0.77,0.82)	0.91 (0.89,0.93)
5 (Deprived)		67.0 (66.4,67.6)	0.70 (0.68,0.73)	0.86 (0.84,0.88)
Presence of self-reported chronic medical condition				
No Long Term Condition		66.3 (65.8,66.7)	Ref	Ref
Long term Condition		76.8 (76.4,77.1)	1.68 (1.66,1.70)	1.29 (1.27,1.30)
Presence of self-reported long-standing psychological or emotional condition				
No Mental Health Condition		71.9 (71.5,72.2)	Ref	Ref
Mental Health Condition		75.3 (74.8,75.8)	1.19 (1.17,1.22)	1.25 (1.22,1.27)

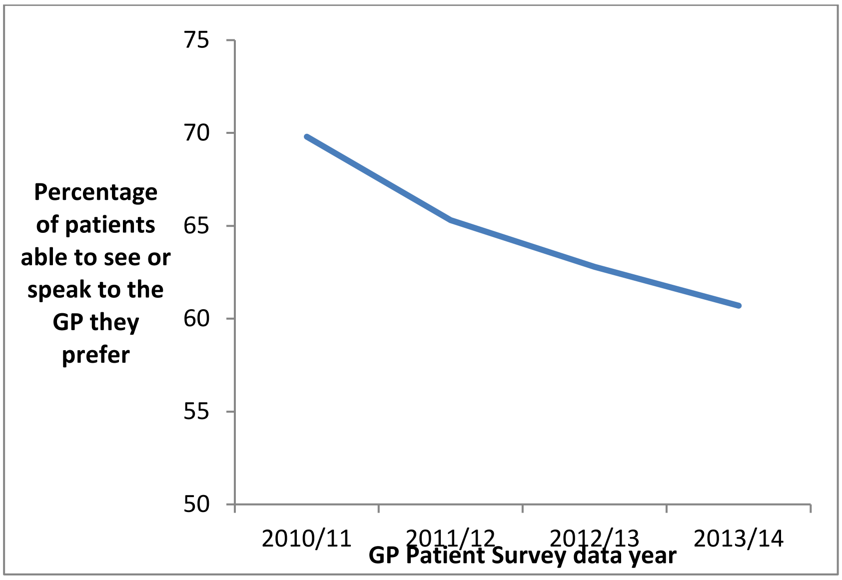
[continued on next page]

	Weighted prevalence* (95% CI)	Crude Weighted ORs* (95% CI)	Adjusted ORs** (95% CI)
Number of practice GPs			
1GP	78.5 (77.3,79.6)	1.10 (1.01,1.19)	1.42 (1.33,1.52)
2GPs	76.9 (76.0,77.7)	Ref	Ref
3GPs	74.5 (73.6,75.4)	0.88 (0.82,0.94)	0.76 (0.71,0.81)
4GPs	73.4 (72.6,74.3)	0.83 (0.78,0.89)	0.66 (0.61,0.70)
5GPs	72.0 (71.1,72.9)	0.77 (0.73,0.83)	0.57 (0.53,0.61)
6-9GPs	69.7 (69.2,70.3)	0.69 (0.66,0.73)	0.48 (0.45,0.51)
10+ GPS	68.8 (67.0,70.7)	0.67 (0.60,0.74)	0.44 (0.40,0.49)
Type of appointments sought in previous 6 months* **			
No appointments requested	73.6 (73.2,74.0)	1.23 (1.21,1.25)	1.17 (1.15,1.19)
Urgent only	69.4 (68.9,69.8)	Ref	Ref
Non-urgent only	78.8 (78.4,79.2)	1.64 (1.60,1.68)	1.59 (1.57,1.62)
Both urgent and non-urgent	70.3 (69.9,70.7)	1.05 (1.03,1.06)	1.12 (1.10,1.13)

* Estimated from weighted unadjusted analysis. P<0.001 for all association (joint tests for categorical variables).
** Estimated from a single multivariate logistic regression model including all variables which appear in the table plus a random practice intercept . P<0.001 for all associations. (joint tests for categorical variables).
*** Based on patients’ response to questions 6 and 9 (see Methods)

Analysis 2 - Seeing the clinician of the patient’s choice: changes over time

Figure 9 shows the percentage of people who have a preference to see a particular doctor who were actually able to do so last time they had a consultation. This shows that the percentage of patients able to see the GP of their choice has declined year on year for the past four years form 70% in 2010/11 to 61% in 2013/14.



Note: Percentages presented have been weighted for survey design and non-response such that they represent the national population rather than respondents.

Figure 9. Percentage of people able to see or speak to the GP they prefer 'a lot of the time', 'almost always' or 'always' (of people who say they have a particular GP they prefer to see – data from GP Patient Survey)

Analysis 3 - Seeing the clinician of the patient's choice: association with subsequent rating

Of the 903,357 people who responded to the 2013/14 GP patient survey 870,085 answered the question regarding what they wanted to do last time they wanted to see or speak to someone from their GP surgery. Table 7 shows how that question was answered. Accounting for non-response and design weighting suggests that over three quarters of patients only want to see or speak to a GP, whilst just under 15% only want to see or speak to a nurse.

Table 7. Descriptive statistics showing what patients wanted to do last time they contacted the GP surgery.

Last time you wanted to see or speak to a GP or nurse from your GP surgery: What did you want to do?	n	Weighted %
Wanted to see and/or speak to a GP	653,526	77.7%
Wanted to see and/or speak to a Nurse	139,300	14.5%
Wanted a home visit	12,873	1.2%
Didn't mind/wasn't sure	15,404	2.4%
One or more of the above	48,982	4.3%
Total	870,085	100.0%

We then compared the type of appointment patients wanted and the type of appointment they actually received (Table 8). By and large the vast majority of people get what they want, with 96% of people who want to see or speak to a GP achieving this compared with 92% of people who want to see or speak to a nurse (Table 8). The number wanting a home visit who received one was lower at 80%.

Table 8. Association between what type of appointment patients wanted and what they got. The Table shows weighted (for design and non-response) percentages of what people got for each category of what they wanted.

	What type of appointment did you get?				
Last time you wanted to see or speak to a GP or nurse from your GP surgery: What did you want to do?	To see and/or speak to a GP	To see and/or speak to a Nurse	A home visit	One or more of the above	Total
Wanted to see and/or speak to a GP	95.9%	2.9%	0.1%	1.1%	100.0%
Wanted to see and/or speak to a Nurse	6.3%	92.2%	0.1%	1.4%	100.0%
Wanted a home visit	15.7%	2.2%	79.6%	2.5%	100.0%
Didn't mind/wasn't sure	67.1%	29.1%	1.1%	2.7%	100.0%
One or more of the above	29.7%	8.6%	0.7%	61.1%	100.0%
Overall	77.6%	17.5%	1.1%	3.8%	100.0%

The results of the analysis investigating the association between reported nurse-patient communication and the mismatch between what people wanted and what they got in terms of appointment is shown in Table 9. This analysis is restricted to the 121,086 patients who reported seeing or speaking to a nurse on their last visit to the GP surgery, who also have complete information on the covariates used in the adjusted model. The mean communication score for those who wanted to speak to a nurse and did the same was 90.0 out of 100. For all other combinations nurse-patient communication scores were, on average, lower ($p<0.001$). This difference was largest for those wanting to see or speak to a GP who then saw or spoke to a nurse, with the lowest scores for those who wanted to see a GP but spoke to a nurse (adjusted difference compared to those who wanted to see a nurse and did see a nurse=-10.5 95% CI -11.7, -9.2).

Table 9. Results of the regression analysis examining the association between nurse-patient communication scores and what patients wanted to do and what they did do among those who saw a nurse at their last visit to the GP surgery.

Wanted to happen	**Actually happened**	**N (%)**	**Mean Communication Score**	**Crude difference* (95% CI)**	**Adjusted difference** (95% CI)**
See a nurse	Saw a nurse	105140 (86.8%)	90.0	Reference	Reference
See a nurse	Spoke to a nurse	517 (0.4%)	88.3	-1.70 (-2.97, -0.44)	-1.40 (-2.63, -0.18)
Speak to a nurse	Saw a nurse	1170 (1.0%)	87.6	-2.37 (-3.21, -1.53)	-2.05 (-2.87, -1.23)
Speak to a nurse	Spoke to a nurse	1697 (1.4%)	88.9	-1.09 (-1.79, -0.38)	-0.93 (-1.61, -0.24)
See a GP	Saw a nurse	10916 (9.0%)	82.7	-7.31 (-7.60, -7.02)	-5.99 (-6.28, -5.71)
See a GP	Spoke to a nurse	538 (0.4%)	77.8	-12.16 (-13.40, -10.92)	-10.47 (-11.68, -9.27)
Speak to a GP	Saw a nurse	819 (0.7%)	85.7	-4.25 (-5.26, -3.25)	-3.35 (-4.33, -2.38)
Speak to a GP	Spoke to a nurse	289 (0.2%)	80.2	-9.80 (-11.49, -8.11)	-8.66 (-10.30, -7.03)

*p<0.001 (joint test)

**Also adjusted for age, gender, ethnicity, confidence in managing their own health, the presence of a long standing heath condition, deprivation (fixed effects) and practice (random effect) p<0.001 (joint test)

The parallel analysis for those patients who saw or spoke to a GP is shown in Table 10. This analysis is restricted the 497,302 patients who reported seeing or speaking to a doctor on their last visit to the GP surgery and also have complete information to the covariates used in the adjusted model. The mean communication score for those who wanted to speak to a doctor and did so was 85.4 out of 100. For the majority of the remaining categories doctor-patient communication scores were, on average, lower (p<0.001). The differences were small in most cases (1 to 2 points), though they were greatest when the patient wanted to see a GP and ended up speaking to either a GP or a nurse on the phone.

Table 10. Results of the regression analysis examining the association between doctor-patient communication scores and what patients wanted to do and what they did do among those who saw a doctor at their last visit to the GP surgery.

Wanted to happen	Actually happened	N (%)	Mean Communication Score	Crude difference* (95% CI)	Adjusted difference** (95% CI)
See a GP	Saw a GP	450555 (90.6%)	85.4	Reference	Reference
See a GP	Spoke to a GP	9127 (1.8%)	80.0	-5.41 (-5.77, -5.05)	-4.49 (-4.84, -4.14)
Speak to a GP	Saw a GP	8281 (1.7%)	85.2	-0.21 (-0.59, 0.17)	-0.23 (-0.59, 0.13)
Speak to a GP	Spoke to a GP	22039 (4.4%)	86.9	1.54 (1.30, 1.77)	0.96 (0.73, 1.19)
See a nurse	Saw a GP	5831 (1.2%)	84.5	-0.85 (-1.30, -0.40)	-1.22 (-1.65, -0.80)
See a nurse	Spoke to a GP	378 (0.1%)	82.0	-3.41 (-5.17, -1.64)	-3.76 (-5.43, -2.10)
Speak to a nurse	Saw a GP	913 (0.2%)	84.3	-1.07 (-2.21, 0.07)	-0.94 (-2.01, 0.13)
Speak to a nurse	Spoke to a GP	178 (0.0%)	83.7	-1.64 (-4.21, 0.93)	-2.16 (-4.59, 0.26)

*p<0.001 (joint test)

**Also adjusted for age, gender, ethnicity, confidence in managing their own health, the presence of a long standing heath condition, deprivation (fixed effects) and practice (random effect) p<0.001 (joint test)

Conclusions

Our analyses show that most patients have a particular GP whom they prefer to see. It is sometimes suggested that this only matters for some population groups (e.g. not for young people) but we found this is not the case. Even among 18 to 24 year olds, more than 50% of respondents to the GP Patient Survey have a particular doctor they prefer to see, rising to over 80% in people over 75. Disturbingly, a large percentage of people who have such a preference are unable to see the doctor of their choice. This percentage has risen from 30% to 39% over just the past five years. We can only speculate on the reasons for this, and the change is likely to be due to a range of factors including the pressure on GPs to increase access by offering same day appointments and by opening for longer hours or on more days. In addition the increasing proportion of GPs working part time may make it more difficult for patients to see the GP of their choice.

One of the criticisms of patient surveys are the very positive scores which patients give, scores which may not represent the totality of their experience (as we have shown in chapters 2 and 3). However, we do see less positive scores for ratings of being able to see a doctor of your choice, with 40% of patients responding to the GP Patient Survey now saying they are unable regularly to see the doctor of their choice. This is clearly an important quality issue for the NHS which has received scant attention from governments which remain focused on access. Providing good continuity is difficult in the context of contemporary general practice, but there are ways of organising practice to increase patients' ability to choose the doctor they see. The Royal College of General Practitioners has published a toolkit on the subject [135] and we have also published guidelines on how practices can improve the continuity of care they provide. [136] This is certainly an area which deserves more priority in the NHS.

We are able to get some insight into the impact of this on patient experience from our analysis of data from patients who have seen a nurse where they had originally wanted to see a doctor. These patients report substantially worse experience with their subsequent consultation with a nurse, and we have no reason to think that the nurses were behaving differently to these patients from any others (and nurses generally get very high scores for their communication with patients).

SECTION B

Understanding patient experience in minority ethnic groups

Chapter 5. Analyses of GP Patient Survey data to explore variations in patient experience by ethnic group and practice

Abstract

Background

Within the UK, there is particular concern over South Asian patients’ experience of care, with consistently more negative ratings across a wide range of measures. The nature and potential drivers of the reported variations in care in South Asian groups has yet to be fully explored. In this workstream, we aimed to investigate a number of potentially contributory factors to variations in communication with primary care professionals related to ethnicity and practice.

Methods

Analyses of GP Patient Survey data.

Results

South Asian respondents report more negative experiences of GP-patient communication than their White British counterparts. Around half of this variation may be attributed to the concentration of these patients in low-performing practices. However, the effect of ethnicity on reported GP-communication varies by age and gender, with poorer experience scores being particularly marked in older, female, Asian patients. There was no evidence of differential item functioning of the communication items for White British and South Asian patients. These findings increase the likelihood that there are true differences in the quality of care received by South Asian groups compared with the White British majority. A substantial proportion in the variability in practice scores for GP-patient communication can be explained by practice factors.

Conclusions

Reports of communication with primary care professionals are more negative for South Asian respondents. Whilst practice factors are an important driver of this, even within the same practice South Asian patients (particularly those who are older and female) are likely to experience lower quality of communication.

Introduction and rationale for the study

Systematic variations in experience of health care in relation to ethnicity, age, gender, health, and socioeconomic status have long been documented in the UK. [137-139] In 2014, NHS England reiterated concerns about variations in the quality of primary care for disadvantaged groups, stating that: "People have a right to high quality services, irrespective of who they are, their social status, where they live, or what needs they have." [4] A particular focus has been the experience of some minority ethnic groups, who have reported consistently lower patient experience scores compared to the majority population in both the UK and the US. [68, 140-143] Previous analyses of patient experience data conducted by the authors highlighted that South Asian patients reported particularly negative experiences, including for waiting times for GP appointments, time spent waiting in surgeries for consultations to start, and continuity of care. [68]

Several potential explanations have been proposed for the lower patient experience ratings given by South Asian respondents in response to surveys. Broadly, these relate to whether South Asian patients receive lower quality care, or whether they receive similar care, but rate this more negatively. [68, 73, 144] For example, differences in the use of questionnaire response scales [73] may lead to South Asian groups being less likely to endorse the most positive options when asked to evaluate a doctor's communication skills. Alternatively, there may be systematic variations in evaluations of consultations because South Asian respondents vary in their expectations of, or preferences for, care. Finally, of course, it is possible that reported poorer experiences of care do reflect actual differences in the care received by these patient groups. In this workstream, we set out to explore in more detail the nature and potential drivers of the reported variations in care in South Asian groups, using existing GP Patient Survey data. Experimental work to explore how South Asian and White British participants rate simulated consultations is detailed in Chapter 6.

Structure of work package

We undertook a series of analyses of GP Patient Survey data to investigate variations in patient experience for South Asian groups. This work was undertaken across four workstreams:

1. An exploration of whether the low scores of minority ethnic and other socio-demographic groups reflect their concentration in poorly performing primary care practices.
2. Building on the above, further analyses to determine how reported GP-patient communication varies between patients from different ethnic groups, stratified by age and gender.
3. An analysis, using item response theory, to test for evidence as to whether the GP Patient Survey communication items perform differently for South Asian and White British respondents.
4. Finally, in addition to the above patient-level analyses, we explored how differences between practices influence GP-patient communication scores.

Changes to study methods from original protocol

The aim of this workstream, as stated in the original protocol, was:

"To understand the reasons why minority ethnic groups, especially South Asians, give lower scores on patient surveys compared to the White British population (aim 5)."

We conducted all analyses outlined in the original protocol. However, part (2), an exploration of how reported GP-patient communication varies by ethnicity stratified by age and gender, was an additional analysis undertaken to gain better insight into the particular combinations of patient characteristics associated with the most negative reported experiences of care.

Background to the GP Patient Survey

The GP Patient Survey was started in 2007 as a national postal survey of primary care patients. Each year, it takes a random sample of patients registered at all NHS primary care practices and sends out a questionnaire covering key aspects of patient experience, including access, waiting times, and communication with doctors and nurses. Findings from the survey are disseminated widely and are available to practitioners and patients via the dedicated GP Patient Survey website (www.gp-patient.co.uk). In the most recent year, 2014/2015, a questionnaire was sent to 2.6 million patients of whom 858,381 responded (32.5% response rate). Respondents may complete the survey via post or online, including in British Sign Language, and in 13 languages other than English either online or via telephone.

The original GP Patient Survey questionnaire was developed iteratively, with guidance from stakeholders and experts, cognitive testing of items, and extensive piloting. [119] It has been further developed over the years, with changes to the content and technical aspects including survey weighting.

Workstream 1. Do poor patient experience scores of minority ethnic groups reflect their concentration in poorly performing primary care practices?

Aims and objectives

The aim of this workstream was to investigate the causes of socio-demographic variations in patient experience. There were two specific objectives:

1. Do minority ethnic group differences in reported GP-patient and nurse-patient communication arise from the concentration of minority ethnic patients in practices with lower than average performance?
2. Do minority ethnic group differences in reported GP-patient and nurse-patient communication vary substantially across practices?

Methods

We analysed data from the 2009/10 GP Patient Survey. Drawing on our previous principal components analysis of survey data, we constructed a measure of reported GP-patient and nurse-patient communication from seven communication items (Box 4). [119] From these, we created a composite score for all responders who provided three or more informative responses: this was derived by linear rescaling of the responses between 0 and 100 and taking the mean of all sub-items answered.

Last time you saw or spoke to a GP/nurse from your GP surgery, how good was that GP/nurse at each of the following?

Please put an X in one box for each row

	Very good		Good		Neither good nor poor		Poor		Very poor		Doesn't apply *
Giving you enough time...........................	☐	...	☐	...	☐	...	☐	...	☐	...	☐
Asking about your symptoms	☐	...	☐	...	☐	...	☐	...	☐	...	☐
Listening to you	☐	...	☐	...	☐	...	☐	...	☐	...	☐
Explaining tests and treatments	☐	...	☐	...	☐	...	☐	...	☐	...	☐
Involving you in decisions about your care	☐	...	☐	...	☐	...	☐	...	☐	...	☐
Treating you with care and concern	☐	...	☐	...	☐	...	☐	...	☐	...	☐
Taking your problems seriously	☐	...	☐	...	☐	...	☐	...	☐	...	☐

* Considered to be uninformative for the purposes of our analysis

Box 4. GP and nurse communication items for the 2009/10 GP Patient Survey

Patient reported gender, age group, gender, ethnicity, self-rated health, and presence of a long-standing psychological or emotional condition were taken directly from survey responses. Socio-economic status was measured using an area based approach, the Index of Multiple Deprivation (IMD), based on the patient's residential postcode. [145] For analysis, we split the IMD into five groups, based on national quintiles.

To examine our first objective (to distinguish the effects of the concentration of some minority ethnic groups in low-scoring practices from the variation of scores of different population groups within practices), we combined two analytical strategies:

1. Fixed effects multivariable linear regression models to predict patient experience measures only from patient socio-demographic characteristics. These models estimate overall sociodemographic differences in patient experience which arise both because some patient groups are concentrated in low-performing practices and because the scores of patients of different groups vary within the same practices.

2. Mixed effects models that included patient socio-demographic variables as fixed effects plus a random effect (intercept) for practice. These models estimate only the socio-

demographic differences that arise because the scores of patients of different groups vary within the same practices.

We used the difference between the coefficients of the first and the second models to indicate the amount of overall difference arising from the concentration of any given population group in practices with low scores.

To examine our second objective (to assess whether sociodemographic differences are consistent among practices) we added random effects (slopes) to the above models, corresponding to the interaction of each patient characteristic variable with the ‘practice’ random effect (random slope random intercept models). From those models, using a normal approximation, we derived the ‘95% midrange of practice-level coefficients’ for each socio-demographic group, which indicates the range of practice-level sociodemographic differences within which 95% of all practices lie.

SAS v.9.2 was used for random slope random intercept models and Stata v.11 (StataCorp, Texas, USA) for all other analyses.

Results

There were 2,163,456 responses to the GP Patient Survey in 2009/10, representing an overall response rate of 38%. Table 11 shows the response by ethnic group.

Table 11. Ethnicity of respondents to the 2009/10 GP Patient Survey

	Survey respondents (n)	**Percentage of survey respondents**
Ethnic group		
White		
White British	1,718,133	82.0
Irish	29,930	1.4
Any other white	61,087	2.9
Mixed		
White and Black Caribbean	4,549	0.2
White and Black African	2,825	0.1
White and Asian	4,142	0.2
Any other mixed	3,564	0.2
South Asian		
Indian	53,484	2.6
Pakistani	33,517	1.6
Bangladeshi	10,974	0.5
Any other Asian	14,930	0.7
Black		
Black Caribbean	25,231	1.2
Black African	28,349	1.4
Any other Black	4,174	0.2
Chinese	9,759	0.5
Other ethnic group	90,644	4.3

Reported experiences of GP-patient communication in Bangladeshi, Pakistani, Indian and Chinese respondents were respectively -9, -7, -6 and -8 percentile points more negative than White British patients (Table 12).

There were similar findings for nurse-patient communication (Table 13). Our first objective in this strand of work was to examine whether such overall minority ethnic differences in experiences of care arise from the concentration of these patients in practices with lower than average performance. By comparing the coefficients obtained from the fixed and mixed effects model for GP-patient communication, we identified that the concentration of minority ethnic groups in low-scoring practices was responsible for about 50% of the difference between South Asian and White British patients. However, even after accounting for the effect of the concentration of these groups in practices with lower scores, South Asian patients reported more negative experiences of care than their White British counterparts within the same practice.

Table 12. Socio-demographic differences in reports of GP-patient communication

Ethnic group	**Overall difference *** **Difference (SE)**	**Difference attributable to different evaluation of care within the same practice *** **Difference (SE)**	**Difference attributable to concentration of different patient groups in practices with different mean scores**	**Percentage of overall difference attributable to patient group concentration in practices with different mean scores**
White				
White British	Reference			
Irish	-0.2 (0.141)	0.6 (0.138)	-0.8	353% ††
Any other white	-4.1 (0.096)	-3.2 (0.094)	-0.9	22%
Mixed				
White and Black Caribbean	-1.9 (0.355)	-0.8 † (0.346)	-1.1	56%
White and Black African	-3.5 (0.447)	-1.9 (0.435)	-1.6	46%
White and Asian	-3.4 (0.358)	-2.2 (0.348)	-1.1	33%
Any other mixed	-4.7 (0.405)	-3.3 (0.394)	-1.4	31%
South Asian				
Indian	-6.1 (0.101)	-3.2 (0.109)	-3.0	48%
Pakistani	-7.2 (0.132)	-3.8 (0.145)	-3.4	48%
Bangladeshi	-8.6 (0.233)	-5.3 (0.242)	-3.4	39%
Any other Asian	-4.3 (0.194)	-2.1 (0.192)	-2.2	51%

Black				
Black Caribbean	-2.7 (0.155)	-0.5 (0.156)	-2.2	82%
Black African	-2.6 (0.143)	-0.2 ** (0.144)	-2.4	94%
Any other Black	-2.0 (0.405)	-0.2 ‡ (0.394)	-1.8	89%
Chinese	-8.3 (0.230)	-7.2 (0.225)	-1.1	14%
Other ethnic group	-4.7 (0.081)	-3.2 (0.081)	-1.5	32%

* All coefficients are significant at the <0.001 level except as annotated: †p=0.015; ** p= 0.269; ‡ p=0.579

† Proportions > 100% reflect situations where differences attributable to different evaluations of care within the same practice, and differences attributable to concentration of different patient groups in practices with different mean scores are opposite in direction. Here, for example, Irish White patients are concentrated in low-scoring practices but report better care compared with White British patients looked after by the same practices.

Table 13. Socio-demographic differences in reports of nurse-patient communication (scale 0-100)

Ethnic group	**Overall difference** **Difference (SE)**	**Difference attributable to different evaluation of care within the same practice** **Difference (SE)**	**Difference attributable to concentration of different patient groups in practices with different mean scores**	**Percentage of overall difference attributable to patient group concentration in practices with different mean scores**
White				
White British	Reference			
Irish	-0.5 (0.168)	0.4 (0.166)	-0.9	>+/- 100% *
Any other white	-3.2 (0.118)	-2.4 (0.117)	-0.8	25%
Mixed				
White and Black Caribbean	-1.7 (0.446)	-0.8 (0.439)	-0.9	51%
White and Black African	-4.0 (0.570)	-2.5 (0.561)	-1.5	38%
White and Asian	-4.2 (0.459)	-3.3 (0.452)	-0.9	22%
Any other mixed	-4.1 (0.507)	-2.6 (0.499)	-1.5	37%
South Asian				
Indian	-7.2 (0.123)	-5.1 (0.134)	-2.1	29%
Pakistani	-7.8 (0.165)	-5.9 (0.179)	-1.9	24%
Bangladeshi	-9.6 (0.309)	-7.3 (0.319)	-2.3	24%
Any other Asian	-6.2 (0.244)	-4.3 (0.244)	-1.9	30%

Black				
Black Caribbean	-3.4 (0.192)	-1.4 (0.195)	-2.0	60%
Black African	-4.0 (0.1481)	-1.4 (0.184)	-2.2	55%
Any other Black	-3.6 (0.492)	-1.9 (0.485)	-1.7	47%
Chinese	-9.4 (0.314)	-8.3 (0.311)	-1.2	12%
Other ethnic group	-4.7 (0.101)	-3.3 (0.102)	-1.4	29%

*Proportions >100% reflect situations where differences attributable to different evaluation of care within the same practice, and differences attributable to concentration of different patient groups in practices with different mean scores are opposite in direction.

Our second objective was to examine whether minority ethnic differences varied between as well as within practices. Table 14 shows that within-practice ethnic group differences in reported GP-patient communication varied substantially across practices, alongside other key measures of patient experience. On average, South Asian patients evaluated doctor communication more negatively than White British patients (-4 percentile points): however, in some practices South Asian patients reported more positive experiences of GP-patient communication that their White British counterparts (95% practice midrange for differences in doctor communication: -13 to +4 percentile points (positive values indicate better patient experience than the majority White British group). Again, we found a similar picture for nurse-patient communication (Table 15).

Table 14. Mean ethnic group difference (percentile points) and degree of consistency in ethnic group differences across practices (indicated by the respective 95% midrange) for doctor-patient communication ratings

Ethnic group	**Mean difference in doctor-patient communication scores †**	**95% midrange of practice differences**	
		Lower limit	Upper limit
South Asian *	-4.3	-12.6	4.0
Black *	-1.4	-7.9	5.0
Chinese *	-8.5	-18.3	1.3
Mixed *	-3.9	-16.1	8.2
Other *	-4.3	-11.7	3.1

[*] All interaction (*case mix adjuster * practice)* variance components were significant <0.0001

† All differences <0.0001

Table 15. Mean ethnic group difference (percentile points) and degree of consistency in ethnic group differences across practices (indicated by the respective 95% midrange) for nurse-patient communication ratings

Ethnic group	**Mean difference in nurse-patient communication scores †**	**95% midrange of practice differences**	
		Lower limit	Upper limit
South Asian *	-5.9	-14.0	2.1
Black *	-2.2	-8.3	3.8
Chinese *	-9.2	-23.6	5.2
Mixed *	-3.4	-18.2	11.5
Other *	-3.9	-12.0	4.3

* All interaction (*case mix adjuster * practice)* variance components were significant <0.0001

† All differences <0.0001

Summary

This analysis of GP Patient Survey data confirmed that South Asian respondents report substantially more negative experiences of patient communication than their White British counterparts. Around half of this was due to the concentration of these patients in low-performing practices. However, differences in reported experiences also varied substantially between practices: as well as more negative reports of care, in some practices South Asian patients evaluated their experience similarly or more positively than their White British counterparts.

These findings suggest a number of drivers that may be behind more negative reports of GP-patient communication seen in national patient experience surveys. However, the experimental vignette work we conducted (see Chapter 5) enabled us to determine more clearly where the most plausible explanations lie.

Workstream 2. How does reported GP-patient communication vary between patients from different ethnic groups, stratified by age and gender?

This material is reproduced from, Burt J, Lloyd C, Campbell J, Roland, M and Abel, G. Variations in GP-patient communication by ethnicity, age, and gender: evidence from a national primary care patient survey. Br J Gen Pract 2016; DOI: 10.3399/bjgp15X687637[176] *under Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.*

Aims

Whilst our earlier analyses confirmed the variation in reported experience between minority ethnic groups within the GP Patient Survey, the question remained as to whether more negative experiences of care are consistent across respondents within a particular ethnic group. Recently, interactions between age and ethnicity have been identified for patient reports of the number of GP consultations which take place before hospital referral for cancer. [146] To explore whether such interactions exist for our focus area of patient experience, GP-patient communication, we undertook further analysis of GP Patient Survey data to determine how reported GP-patient communication varies between patients from different ethnic groups by age and gender.

Methods

We analysed data from the 2012/13 and 2013/14 GP Patient Survey: by combining data from two years of the survey, we were able to increase the number of responses available for analysis from small ethnic groups. No patient receives the survey in two consecutive years, so there is no risk of double counting respondents.

Following the same approach outlined earlier in this section, we constructed a measure of reported GP-patient communication from the following five communication items used in the most recent GP Patient Surveys (see Box 5: these were reduced from the seven used in earlier questionnaires). [119]

Last time you saw or spoke to a GP/nurse from your GP surgery, how good was that GP/nurse at each of the following?
Please put an X in one box for each row

	Very good		Good		Neither good nor poor		Poor		Very poor		Doesn’t apply *
Giving you enough time………………………	☐	…	☐	…	☐	…	☐	…	☐	…	☐
Listening to you …………………………………	☐	…	☐	…	☐	…	☐	…	☐	…	☐
Explaining tests and treatments ………….	☐	…	☐	…	☐	…	☐	…	☐	…	☐
Involving you in decisions about your care	☐	…	☐	…	☐	…	☐	…	☐	…	☐
Treating you with care and concern ……….	☐	…	☐	…	☐	…	☐	…	☐	…	☐

* Considered to be uninformative for the purposes of our analysis

Box 5. GP and nurse communication items for the 2012/13 and 2013/14 GP Patient Surveys

We created a composite score for all responders who provided three or more informative responses: this was derived by linear rescaling of the responses between 0 and 100 and taking the mean of all sub-items answered. Patient reported age group, gender and ethnicity were taken directly from survey responses. Health related quality of life was measured using responses to five questions which make up the EuroQol EQ-5D-3L descriptive system. [147] Socio-economic status was measured using an area based approach, the Index of Multiple Deprivation (IMD), based on the patient’s residential postcode. [145] For analysis, we split the IMD into five groups, based on national quintiles.

We used a mixed effect linear regression model with GP-patient communication score as the outcome. The model included age, gender, ethnicity, EQ-5D, and deprivation as fixed effects, as well as a random effect (intercept) for practice to account for the fact that certain patient groups cluster in practices that may perform better or worse overall. We included in the model all possible 2-way interactions between age, gender and ethnicity, as well as the 3-way interaction between them, to allow the effect of ethnicity to vary between different age and gender groups. We used Wald tests of the interaction terms to assess evidence supporting this variation. We then used the models to estimate age and gender specific differences between White British patients and patients of the same age and gender from each of the other ethnic groups. All analyses were carried out using Stata v13.1 (StataCorp, Texas, USA).

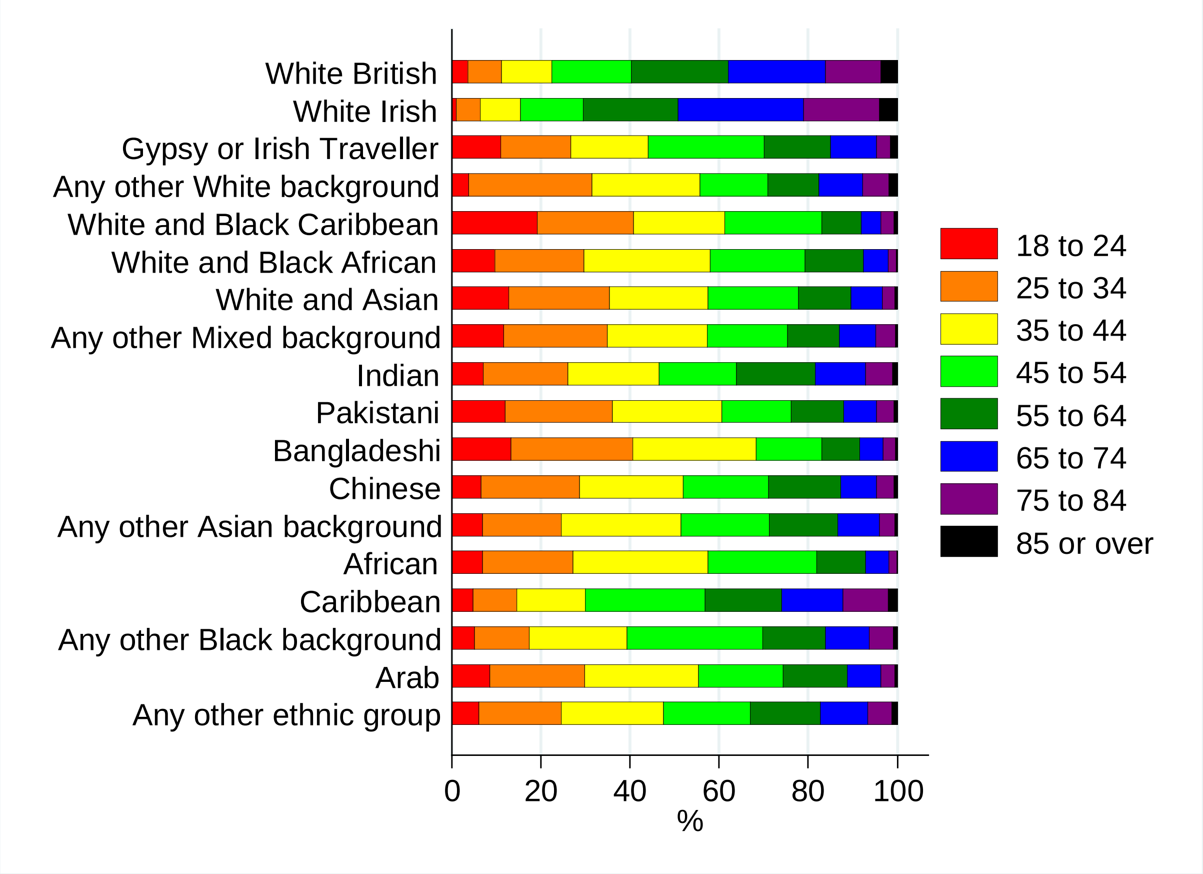
Results

There were 1,874,589 responses to the GP Patient Survey across 2012/13 and 2013/14, representing an overall response rate of 35%. Of these responses, 1,599,801 (85%) had complete data for all items included in our analysis. Table 16 shows the numbers of respondents in each ethnicity group. The largest group of responders were White British (n=1,708,290, 82%), although there were at least 1,800 responders in all but one group (that of Gypsy or Irish Traveller). Figure 9 shows the age composition of each ethnic group. White British and White Irish responders tended to be older than those from other ethnic groups, and are dominated by those aged 55 and over. In contrast, for nearly all other ethnicities the majority of responders were below the age of 45. We therefore had very few responses in the oldest age groups (particularly 85 and over) for a number of ethnicities (see Table 16 for details).

Table 16. Ethnicity make-up of sample for all ages and those aged 85 and over.

Ethnicity		**All ages**		**85 and over**	
		n	**%**	**n**	**%**
White	British	1,323,621	82.7%	49,891	93.1%
	Irish	16,330	1.0%	662	1.2%
	Gypsy or Irish Traveller	401	0.0%	6	0.0%
	Any other white	71,105	4.4%	1,386	2.6%
Mixed / multiple ethnic groups	White and Black Caribbean	3,413	0.2%	26	0.1%
	White and Black African	1,865	0.1%	4	0.0%
	White and Asian	3,171	0.2%	18	0.0%
	Any other mixed	3,340	0.2%	15	0.0%
Asian / Asian British	Indian	38,705	2.4%	425	0.8%
	Pakistani	20,729	1.3%	143	0.3%
	Bangladeshi	6,699	0.4%	23	0.0%
	Chinese	7,986	0.5%	66	0.1%
	Any other Asian	19,812	1.2%	105	0.2%
Black / African / Caribbean / Black British	African	21,131	1.3%	24	0.0%
	Caribbean	13,715	0.9%	275	0.5%
	Any other Black	6,061	0.4%	52	0.1%
Other ethnic group	Arab	2,786	0.2%	16	0.0%
	Other	38,931	2.4%	458	0.9%
Total		1,599,801	100.0%	53,595	100.0%

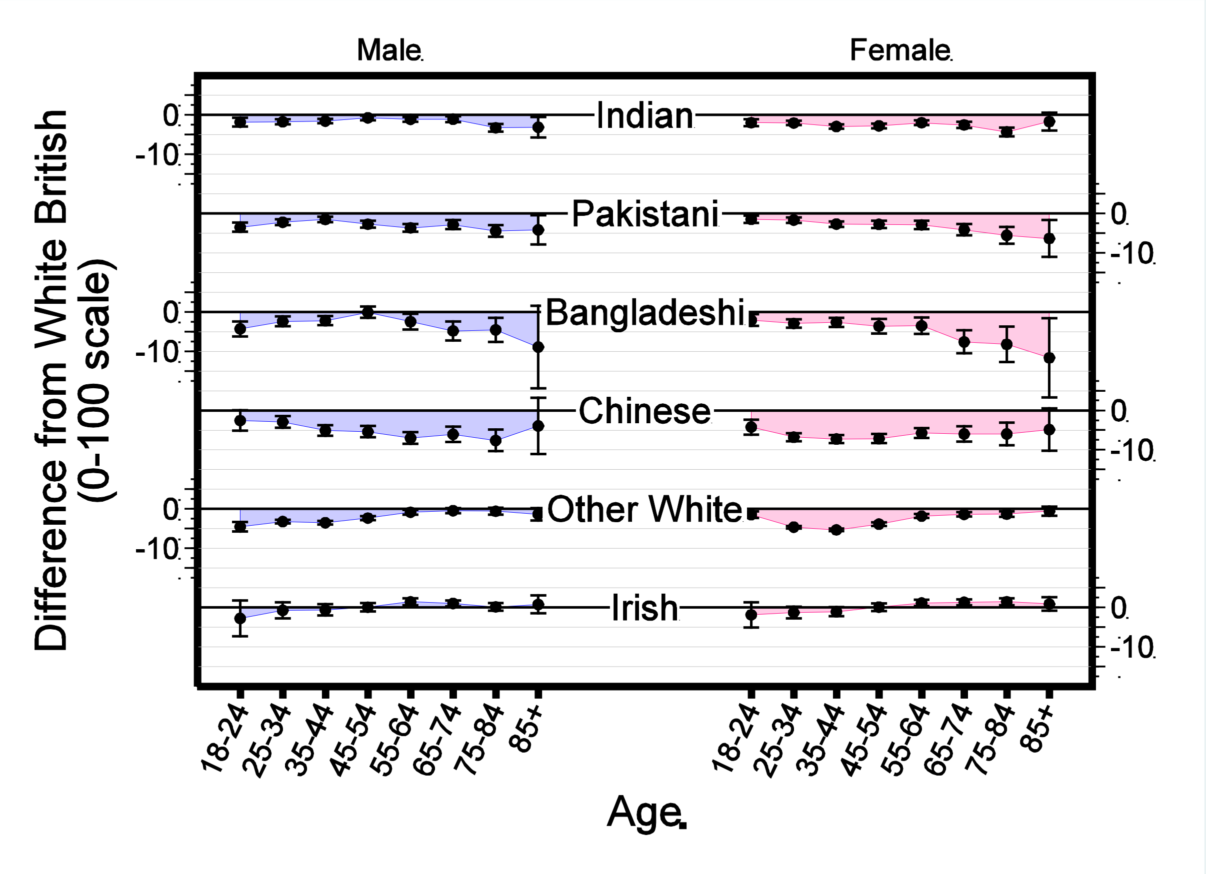
Reproduced from, Burt J, Lloyd C, Campbell J, Roland, M and Abel, G (2016) [176]



Reproduced from, Burt J, Lloyd C, Campbell J, Roland, M and Abel, G (2016) [176]

Figure 10. Age composition of responders according to self-reported ethnicity

From the regression model (adjusting for deprivation, EQ-5D and practice) there was strong evidence ($p<0.001$ for age by gender by ethnicity three-way interaction term) that the effect of ethnicity on reported GP-patient communication varied by both age and gender. [176] Figure 10 shows the age and gender specific adjusted differences between White British responders and responders of the same age and gender from all Asian sub-groups and White (non-British) ethnic groups: negative differences indicate responders reported worse experience than their White British counterparts (i.e. of the same age and gender). Again, as with our previous analyses, the largest differences are seen in Asian ethnic groups, alongside White (non-British) groups.



Reproduced from, Burt J, Lloyd C, Campbell J, Roland, M and Abel, G (2016) [176]

Figure 11. Age and gender specific differences, with 95% confidence intervals, in reported GP-patient communication scores (0-100 scale) between White British patients and responders in Asian and White ethnic groups.

Differences in reported experience of GP-patient communication between Asian groups and the White British group were particularly large for older responders (those aged 55 years and above). This differential effect of ethnicity was particularly marked in Bangladeshi responders, and for women (again see Figure 10 for details). For example, the difference in reported experience scores between a White British 75-84 year old woman and a Bangladeshi woman of the same age was -8.23 points on 0-100 scale (95% CI -12.76, -3.69). However, for Indian, Pakistani and Bangladeshi groups the differences in younger age groups compared to White British responders were fairly small. For example, the difference in reported experience score between a White British 35-44 year old woman and a Pakistani woman of the same age was -2.72 points (95% CI -3.42, -2.02). For Chinese responders, substantial negative differences compared to White British counterparts were seen across all age groups.

In contrast to Asian responders, for those responders identified as ‘Any other White’, ethnic variations in reported communication were largest for younger responders (those aged under 55 years). We found few differences in reported experience at all ages for African, Caribbean and Other Black responders. Due to the smaller sample sizes, our ability to detect differences for mixed ethnic groups is limited. However, we note that there were more substantial (and statistically significant) negative differences for Other Asian women (at all ages), and for White and Asian women (particularly at older ages).

Summary

This analysis of GPPS data has shown that the effect of ethnicity on reported GP-patient communication varies by age and gender. In comparison to White British responders of the same age and gender, poorer experience scores for GP-patient communication are particularly marked in older, female, Asian patients, and in younger ‘Any Other White’ patients. This highlights the need to focus not just on ethnic background, but on how this interacts with other patient characteristics such as age and gender in its association with more negative reported experiences of care.

Workstream 3. Is there evidence that the GP Patient Survey communication items perform differently for South Asian and White British respondents?

This material is reproduced, with permission, from Setodji CM, Elliott MN, Abel G et al. Evaluating Differential Item Functioning in the English General Practice Patient Survey. Comparison of South Asian and White British Subgroups. Medical Care 2015; 53(9): 809-817.[177]

Aims and objectives

As already outlined, observations of poorer reported experience for certain minority ethnic groups may be attributed either to variations in the way in which they rate their care or to variations in the care actually received. Item response theory is one approach to exploring whether observed differences in survey responses may be attributable to true differences in health care or to differences in responses. [148, 149] The aim of this strand of work was to use Item Response Theory modelling to test for evidence that the GP Patient Survey communication items perform differently for South Asian and White British respondents, after controlling for other sociodemographic characteristics.

Methods

We analysed data from the 2011/12 GP Patient Survey. We restricted the analysis to patients who responded to items about experience with GP and nurse care, and who self-reported White British (n = 818,219) or South Asian ethnicity (n = 54,832). As before, we used the five GP-patient communication items (giving enough time; listening; explaining tests and treatments; involving in decisions about care; treating with care and concern), and these same items repeated for nurse-patient communication.

Item Response Theory (IRT) approaches were used to test for differential item functioning (DIF) for White British and South Asian responses (that is, whether White British and South Asian patients have different understanding and scaling of the survey items). We conducted separate analyses for the GP and nurse communication items. In IRT models, items vary in “difficulty” (the extent to which they are easy for providers to “pass”) and patients differ in “ability” (true health care experiences). IRT models also allow subgroups, such as ethnic subgroups, to differ in true experiences or scale use in a way that is *uniform* across all items that attempt to measure a single construct (such as patient experience). Differences between groups, also known as DIF, [149-151] provide evidence that items are not equivalent in meaning across subgroups and an unmeasured dimension other than the intended construct may be influencing item responses. Ideally, IRT models can rely on an unimpeachable anchor item [152] that measures the same construct as the other items but which is known to be completely unaffected by factors other than true care. This is quite rare in practice, so the all-items method also known as the Wald-2 equating method [153, 154] is more commonly used, where designated anchors are not required. This approach links the metric of the construct of interest (patient experience) across South Asian and White British patients simultaneously and then all item parameters (item difficulty and ability) are estimated using the linked construct but they are free to vary between groups, effectively allowing the assessment of whether the differences between the groups that are being compared are consistent across items. In this approach, inconsistent differences across items are taken as evidence of DIF.

The absence of evidence of DIF is not *conclusive* evidence of equivalence, as it may reflect lack of power or it may reflect differences in scale use or expectations that are uniform across items in a scale. Because it is often considered unlikely that scale use and expectations would have the same effects on different items, lack of evidence of DIF in a well-powered study such as the present study, is often seen as suggestive that true differences play a non-negligible role in observed differences in mean scores.

Uni-dimensionality and DIF Analysis

Because of the large sample sizes (818,219 White British and 54,832 South Asians), power to detect statistical significance for even very small differences with a classic χ^2 or Wald test is very high, even after a Benjamini-Hochberg adjustment for multiple comparisons. [154] Consequently, the RMSEA, a transformed Wald χ^2 statistic that measures the degree of misfit independently of sample size, was used for DIF inference. [155] The IRT DIF analyses were conducted in flexMIRT. [154] To assess the possibility that our inferences of White British/South Asian differences were biased by potential confounders such as age, sex, chronic conditions, and quality of life we conducted an additional sensitivity analysis with a matched sample of 54,484 South Asian and 54,484 White British patients with exactly the same characteristics on these potential confounders. The few South Asians with no match (0.63%) were dropped from the analysis.

Results

Using the DIF all-other anchor selection method with Wald-2 equating algorithm, we found no item with DIF. These results are shown in Table 17 for both the full and the matched sample of patients. The RMSEA fit statistic was <0.0085 for all GP items and <0.0140 for all nurse items, suggesting the absence of DIF. In general, discrimination parameters typically range from 0.5 to 2, with higher values indicating items that better discriminate between levels of the latent construct, [156] in this case patient experience. In this study, all item discrimination parameters ("a" in Table 17) exceed 4.4, showing that all items are highly related to the overall score within the GP or nurse item set. The item difficulty parameters, which indicate the level of patient experience θ at which an item has a 50% chance of endorsement, would typically fall between -2 (2.5th percentile) and +2 (97.5th percentile). [157] In this study, they range from -2.62 to 0.03 for GP items (columns b2 to b5 in Table 17) and from -2.80 to -0.01 for nurse items, indicating that the scales best measure average and below-average experiences. They also indicate that a merely average patient experience (θ) results in a 50% chance of endorsing the highest response of "very good," consistent with the high numbers of patients endorsing the "very good" response options.

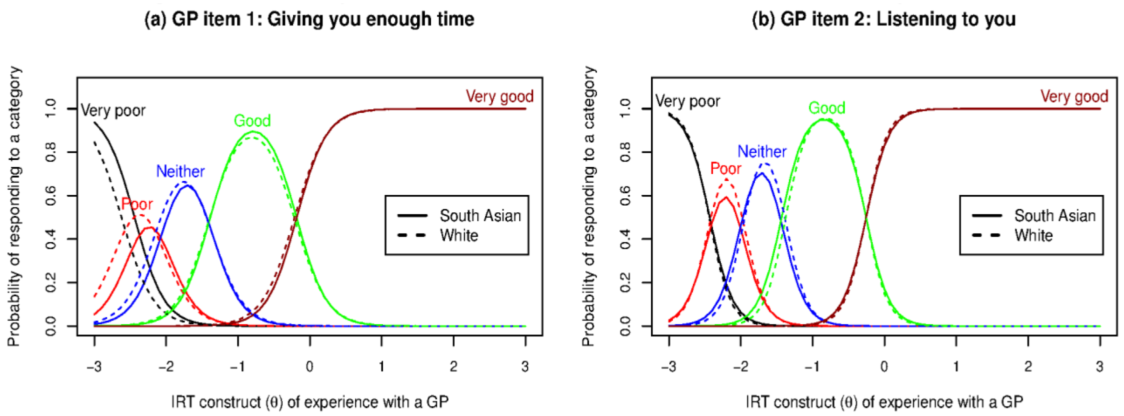
Table 17. Estimated item parameters from graded response model in the full (n = 873,051) and the matched sample (N = 108,968)

Analysis Sample	**Items**	**Ethnicity**	**Full sample**											
			GP items						**Nurses items**					
			Difficulty parameters				***a***	**RMSEA**	**Difficulty parameters**				***a***	**RMSEA**
			b_2	b_3	b_4	b_5			b_2	b_3	b_4	b_5		
Full sample	1. Provider giving you enough time	S Asian	-2.44	-2.03	-1.39	-0.18	4.82	0.0077	-2.65	-2.24	-1.57	-0.28	5.90	0.0084
		White Br	-2.62	-2.11	-1.40	-0.21	4.50		-2.80	-2.33	-1.62	-0.36	5.47	
	2. Provider listening to you	S Asian	-2.43	-1.99	-1.43	-0.25	6.25	0.0068	-2.62	-2.20	-1.56	-0.30	7.86	0.0082
		White Br	-2.44	-1.96	-1.38	-0.26	6.82		-2.66	-2.19	-1.48	-0.30	8.11	
	3. Provider explaining tests and treatments	S Asian	-2.44	-2.02	-1.33	-0.15	5.39	0.0083	-2.63	-2.26	-1.50	-0.26	6.36	0.0098
		White Br	-2.57	-2.08	-1.29	-0.15	5.00		-2.76	-2.30	-1.44	-0.26	5.71	
	4. Provider involving you in decisions about your care	S Asian	-2.34	-1.88	-1.16	0.02	5.15	0.0094	-2.58	-2.15	-1.33	-0.11	5.84	0.0147
		White Br	-2.46	-1.94	-1.11	0.02	4.85		-2.69	-2.20	-1.22	-0.07	5.12	
	5. Provider treating you with care and concern	S Asian	-2.32	-1.91	-1.27	-0.11	5.93	0.0068	-2.56	-2.17	-1.47	-0.22	6.52	0.0069
		White Br	-2.39	-1.93	-1.24	-0.15	6.07		-2.59	-2.18	-1.49	-0.28	6.15	
Matched sample for sensitivity analysis	1. Provider giving you enough time	S Asian	-2.28	-1.87	-1.24	-0.04	4.86	0.0153	-2.54	-2.13	-1.47	-0.20	6.02	0.0184
		White Br	-2.44	-1.94	-1.25	-0.07	4.44		-2.68	-2.20	-1.53	-0.25	5.28	
	2. Provider listening to you	S Asian	-2.27	-1.83	-1.28	-0.12	6.36	0.0083	-2.50	-2.09	-1.45	-0.21	8.04	0.0121
		White Br	-2.28	-1.79	-1.24	-0.12	6.70		-2.54	-2.07	-1.39	-0.22	7.99	
	3. Provider explaining tests and treatments	S Asian	-2.27	-1.86	-1.18	-0.02	5.46	0.0187	-2.52	-2.15	-1.40	-0.17	6.42	0.0207
		White Br	-2.41	-1.91	-1.15	-0.01	4.89		-2.66	-2.18	-1.37	-0.18	5.63	
	4. Provider involving you in decisions about your care	S Asian	-2.18	-1.73	-1.01	0.15	5.15	0.0203	-2.47	-2.04	-1.22	-0.02	5.90	0.0318
		White Br	-2.31	-1.78	-0.97	0.13	4.75		-2.60	-2.10	-1.14	-0.01	5.05	
	5. Provider treating you with care and concern	S Asian	-2.16	-1.75	-1.11	0.03	5.98	0.0108	-2.44	-2.06	-1.37	-0.14	6.63	0.0147
		White Br	-2.21	-1.76	-1.10	-0.01	6.16		-2.44	-2.05	-1.40	-0.20	6.32	

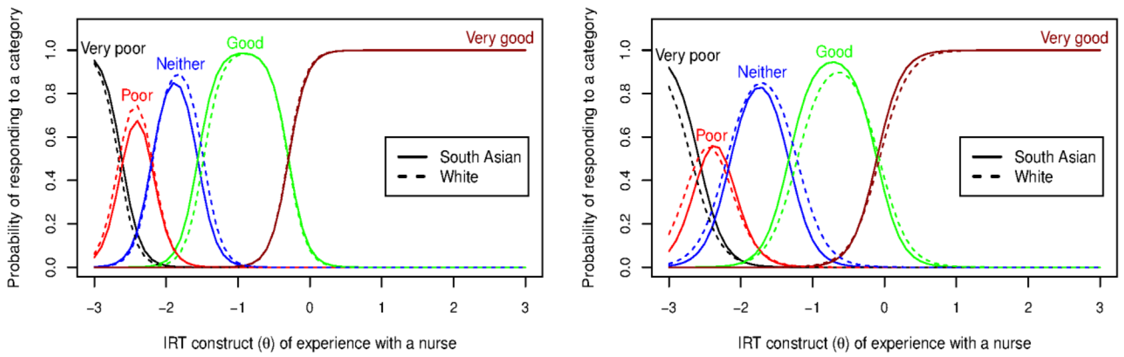
Note: Only observations with at least 1 non-missing item was used in each analyses, so samples of sizes 866,460 and 94,002 were used for full and matched samples for GP items and samples of sizes 783,904 and 82,162 were used for nurse full and matched samples, respectively. a indicates item discrimination parameter; b2, b3, b4, b5, item location or difficulty parameters from the Samejima's graded response model; GP, General Practice; RMSEA, root mean square error approximation; S Asian, South Asian; White Br, White British.

Reproduced, with permission, from Setodji CM, Elliott MN, Abel G et al. (2015)[177]

Figure 12 illustrates the response curves for the parameters in Table 17. For each item, 5 response curves, each representing the probability of endorsing a specific category over the range of the underlying patient experience, showed no visual difference between South Asian and White patients, which is consistent with there being no meaningful DIF. The test characteristic curves depicting the expected scale scores of the GP and nurse items as a function of patient experience on the IRT scale for the 2 groups also shows no difference between the 2 groups (Figure 13).



(a) GP item 1: Giving you enough time **(b) GP item 2: Listening to you**

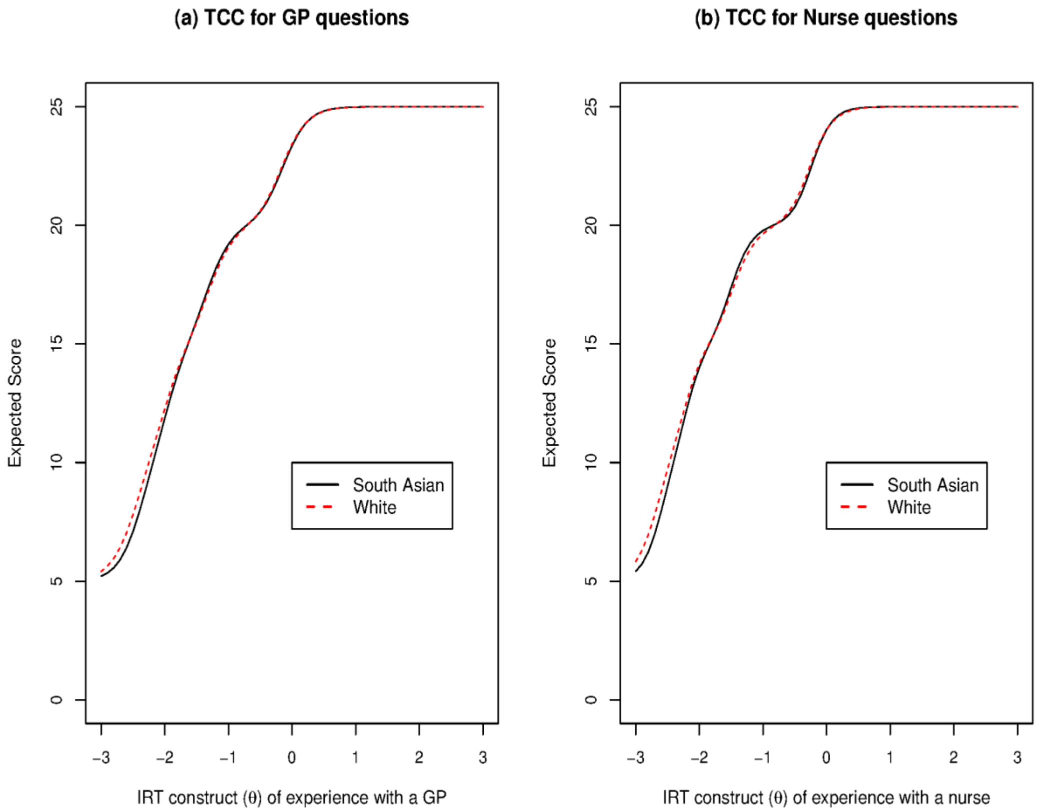


(c) Nurse item 2: Listening to you **(d) Nurse item 4: Involving you in care decisions**

Note: The x-axis is the IRT parameter estimate (γ) for the construct (patient experience) on a z-score metric. The y-axis is the probability of endorsing each response option at given estimates of patient experience quality level for South Asian and White British patients. Samples of 866,460 and 783,904 were used for the GP and the nurse survey items, respectively. Only plot for 2 items are reported, but all the other GP items had similar response curves. (A) GP item 1: giving you enough time; (B) GP item 2: listening to you; (C) Nurse item 2: listening to you; (D) Nurse item 4: involving you in care decisions. GP indicates General Practice; IRT, item response theory.

Reproduced, with permission, from Setodji CM, Elliott MN, Abel G et al. (2015)[177]

Figure 12. Full sample response curves for GP and nurse experience



Reproduced, with permission, from Setodji CM, Elliott MN, Abel G et al. (2015)[177]

Figure 13. Full sample test characteristics curves (TCC). (A) TCC for GP questions; (B) TCC for nurse questions. GP indicates General Practice; IRT, item response theory. [177]

Summary

These analyses found no evidence of meaningful DIF for White British and South Asian patients on GP- and nurse- communication items. These findings remained even after matching patients on a variety of sociodemographic characteristics. We suggest that the lack of evidence of DIF may be consistent either with (1) there being no differences in expectations or scale use between White British and South Asian respondents, or (2) there being differences in expectations and/or scale use between groups that were the same across all items. It is possible that similar differences in scale use may occur across all items, as the response scale and labels remain the same. It is somewhat less likely that there are differences in expectations which remain constant across items which vary in content. Whilst we cannot exclude other possibilities, these findings do increase the likelihood that there is a role for true differences in the quality of care received by South Asian groups in comparison to the White British majority.

Workstream 4. How do practice factors influence GP-patient communication scores?

Aims and objectives

The previous work described in this chapter focussed on patient level factors that may influence reported patient experience, to try and gain insight into what is driving these differences. In Workstream 1, we explored to what extent clustering of certain patient groups in practices with better or worse patient experience scores overall explained differences between groups. Here, we take this one step further and investigate the factors describing a practice which are associated with GP-patient communication scores. We consider three different categories of practice factors; a) practice team, b) practice geography, and c) practice population. In each case we look at the differences in communication scores which are associated with these factors, and how much of the between-practice variance is explained by them.

Methods

We analysed data from the 2009/10 GP Patient Survey. We used the same composite outcome measure for GP-patient communication, using the seven communication items and taking values between 0 and 100. Identical person level exposure variables were used. In addition, we made use of practice level variables from a number of sources.

Practice geography factors

As a proxy for geographic region, we used the former Strategic Health Authority (SHA) to which a practice belonged, of which there are ten. We defined a rurality classification (based on the ONS definitions) according to the postcode of the practice. Both SHA and rurality were included with the GPPS dataset.

Practice professional team factors

Here we used data from the GP census 2009 to calculate for each practice:

a. the number of GPs excluding trainees
b. the number of patients per full time equivalent GP
c. the mean number of years since qualification of the GPs

d. the proportion of male GPs, and
e. the proportion of GPs trained in the UK for their primary medical qualification.

Practice population factors

We calculated a score for socio-economic deprivation for each practice by applying the 2007 Lower Super Output Area Index of Multiple Deprivation proportionately to the practice population. [157] We used GP Patient Survey results to estimate the proportion of Black, Asian, Chinese, mixed race and other non-white patients in each practice. Registered patient numbers, broken down by sex and age group, were provided by the NHS Information Centre and used to calculate the proportion of patients in each practice who were children (under 15 years of age) and the proportion of adult patients in the following age groups: 15-44, 45-64, 65-74, 75-84, 85 and over.

Starting from the random intercept model used in Workstream 1 (including fixed effects for patient age, gender, deprivation, ethnicity and self-rated health, and a random intercept for practice), we added practice level variables for the factors described above. To facilitate comparison between different variables with different distributions and units we scaled all continuous variables (including proportions) such that a difference of one corresponded to the difference between the 95th and 5th percentile of the distribution for that variable. The corresponding coefficients from the regression model can be interpreted as the adjusted differences between practices at either end of the distribution, ignoring outliers. We estimated the amount of variance in practice scores attributable to single or multiple variables by comparing the variance of the random effect in a model containing no practice level variables to a second model containing a single or multiple practice-level variables. We subsequently estimated the amount of variance in practice scores *uniquely* attributable to single or multiple variables by comparing the variance of the random effect in a model containing all practice level variables to a second model with one or more practice level variables omitted.

Results

Table 18 shows the results of the regression model for practice-level variables (patient-level variables are not shown, but are consistent with those shown in Workstream 1). In relation to practice team factors, practices with a large number of GPs, with a high number of patients per FTE, and with doctors who, on average, completed training a longer time ago, tended to have worse GP-communication scores. In comparison, those with a high proportion of GPs trained in the UK had better GP-communication scores. For geographical factors, practices in London and urban areas received the worst scores. Finally, in relation to practice population factors, those practices who served populations with relatively more men, Asian and black patients, and patients under 85, tended to score worse for GP-communication. It is worth noting that these population coefficients have been controlled for individual characteristics, and so do not represent the fact that these patient groups score worse, but rather that practices who have more of these patient groups tend to score worse for all patients.

In total, this model explained 35.4% of the between practice variance in GP-communication scores. Practice team factors explained 25.9% of the total practice-level variance and 11.5% was explained uniquely by practice team factors (i.e. could not be explained by other factors). These numbers compare to 13.8% and 2.7% for practice geography and 18.3% and 3.3% for practice population. Practice team factors are therefore the most important in explaining GP-patient communication scores. Of the practice team factors, the proportion of GPs trained in the UK was the most important, with 5.4% of the total variance uniquely attributable to that variable.

Table 18. Regression coefficients for practice predictors (note models also include individual case-mix variables not shown).

Practice Professional Team	GP Team (95th vs.5th percentile)	Mean Years Since Qualification	-1.5 (-1.9, -1.1)
		Proportion Male GPs	0.3 (0.0, 0.6)
		Proportion UK-Qualified	4.0 (3.6, 4.3)
		Patients per FTE	-1.9 (-2.2, -1.7)
	Number of GPs	1	Ref
		2	-0.6 (-1.0, -0.3)
		3	-0.8 (-1.2, -0.4)
		4	-1.0 (-1.5, -0.6)
		5	-1.1 (-1.5, -0.6)
		6-9	-1.1 (-1.6, -0.7)
		10 or more	-1.4 (-2.0, -0.7)
Practice Geography	SHA (surrogate for region)	North East	3.1 (2.5, 3.7)
		North West	2.9 (2.4, 3.3)
		Yorkshire and the Humber	2.3 (1.8, 2.7)
		East Midlands	1.5 (1.0, 2.0)
		West Midlands	1.6 (1.1, 2.0)
		East of England	1.2 (0.8, 1.7)
		London	Ref
		South East Coast	1.2 (0.7, 1.6)
		South Central	2.0 (1.5, 2.5)
		South West	2.4 (1.9, 2.8)
	Rurality of Practice	Urban >10k Less Sparse	Ref
		Urban >10k Sparse	-0.2 (-2.3, 1.8)
		Town and Fringe	0.1 (-0.3, 0.4)
		Village	2.1 (1.5, 2.6)
		Hamlet & Isolated Dwellings	1.0 (-0.2, 2.3)
Practice Population	Proportion of patients who are: (95th vs.5th percentile)	Female	Ref
		Male	-1.5 (-1.9, -1.2)
		White	Ref
		Mixed	-0.1 (-0.4, 0.3)
		Asian	-1.3 (-1.7, -0.9)
		Black	-0.7 (-1.0, -0.3)
		Chinese	0.0 (-0.1, 0.1)
		Other	0.8 (0.1, 1.4)
		Children <15	-0.7 (-1.1, -0.3)
	Proportion of adult (15 and over) patients who are: (95th vs.5th percentile)	15-44	-2.8 (-3.8, -1.7)
		45-64	Ref
		65-74	-0.7 (-1.6, 0.3)
		75-84	-2.4 (-3.3, -1.6)
		85+	0.3 (-0.3, 0.9)
	Mean Deprivation Quintile	1 (Least Deprived)	0.2 (-0.2, 0.6)
		2	0.2 (-0.2, 0.6)
		3	-0.1 (-0.5, 0.2)
		4	-0.1 (-0.4, 0.3)
		5 (Most Deprived)	Ref

Summary

This analysis demonstrates that a substantial proportion in the variability in practice scores for GP-patient communication can be explained by practice factors. Factors regarding the practice professional team were most important, particularly how many of the GPs were trained in the UK. Where a large proportion of GPs were trained outside of the UK, GP-patient communication scores were substantially lower. Whilst this association might represent the quality of training received in the UK, it is more likely to be a marker of a GP being of a non-British ethnicity. Thus factors which may drive the observed association could plausibly include language and cultural barriers, or discrimination on behalf of the patients. In Workstream 1 (page 112), we demonstrated that around half of the difference between White British and Asian patients was due to their clustering in practices with worse scores overall. This is consistent with our finding that practices in London and other urban areas (where minority ethnic patients would be expected to cluster) and those practices with overseas trained doctors are focussed, have lower scores. Even so, after controlling for these factors, we still found that practices with high proportions of Asian patients had lower GP-communication scores. Interestingly, in other work we have found that South Asian patients who attend a practice where consultations in a concordant South Asian language are offered, the difference between White British and South Asian patients GP-communication ratings decreases. [158]

Overall conclusions

Our analyses of recent GP Patient Survey data have confirmed the presence of substantially more negative experiences of communication by minority ethnic groups, including South Asian groups. A consideration of the interactions between ethnicity, age and gender highlighted that older, female, Asian patients are particularly likely to report negative experiences of communication. Whilst a substantial proportion of these differences may reflect the concentration of such patients in low-performing practices, even within the same practices patients report substantial variations in communication. Our analyses further found no evidence that South Asian and White British groups exhibit differential response tendencies to communication items: whilst experimental work is required to understand whether variations are indeed attributable to poorer quality of care, these findings point to this as the most plausible explanation of identified differences.

South Asian patients may face a number of barriers to high quality care, including poor language proficiency, lack of acculturation and provider-side discrimination. Our analysis of the association of practice factors, particularly the proportion of GPs trained outside of the UK, with reported communication confirms, from a different perspective, the importance of language and cultural factors in determining the quality of communication.

Language is only one part of communication, but an important one. ‘Language-discordance’ occurs when a doctor and patient do not share the same language. The inability to speak English well or at all varies widely between and within ethnic groups: 16.2% of Bangladeshi census respondents, 15.2% of Chinese, 12.2% of ‘Any Other White’ and 11.1% of Pakistani fall into this category. [159] Older Bangladeshi and Pakistani women may be prevented from acquiring English proficiency through family obligations, or cultural and community expectations. [160] A number of studies have suggested that language discordance in clinical encounters may negatively impact on quality of care. [158, 161-163] Challenges in communicating in language discordant consultations can lead to particularly strong tensions between “ideal” standards of communication and what is “good enough”. [164]

Acculturation is concerned with the modification of attitudes or behaviours as people come into contact with a culture other than their own: although its definition and scope are contested, it is frequently used to explain inequalities in health care. [165] Levels of acculturation may lead to variations in perceptions and expectations of providers and care, and ability to navigate the health care system, impacting on reported experience. [166] Previous analysis of patient experience in US primary care for Hispanic patients found no relationship between acculturation levels and patient reports of provider communication, although there was an association with other aspects of patient experience. [166] However, the measurement of acculturation through commonly used language proficiency scales has been criticised for failing to capture its multidimensional nature. [167] Further, a focus on lack of acculturation as a driver of disparities may mask other causal factors, including poverty, the social construction of ethnic identities, and inequities in treatment. [168] Nevertheless, the broad concept of acculturation may be a useful reminder that age, gender and ethnicity groupings could vary in their understanding and navigation of primary care for reasons which are additional to those of language barriers.

Concerns about institutionally ingrained variations in attitudes to patients on the basis of ethnicity have led to a rise in cultural competency training. [169, 170] These approaches have been criticised for placing emphasis on patient characteristics as the drivers of variations in care, rather than on provider- and system-level factors including the potential for stereotyping of or bias towards particular groups. [171] However, our analysis shows that provider- or system-side factors do not occur in reaction to ethnicity alone, but in response to the inter-relationship between ethnicity, gender and age. It is the combination of these factors which may identify groups with particular needs, such as those patients with the lowest levels of English proficiency. We therefore need to focus not just on differences between groups but also on differences within them, considering how ethnicity, gender, age and other categories of social identity interact with each other to create different experiences and outcomes: the study of such interactions has been termed *intersectionality*. [172]

Strengths and limitations

GPPS data are derived from a large, randomly selected sample designed to be representative of patients registered with a practice in England or Wales. [105] However, response rates to GPPS are low: for the years we analysed, these ranged from 34% to 38% though recent reviews suggest response rates are not a strong indicator of non-response bias in surveys which use probability sampling. [173] Unfortunately because ethnicity is not extracted from medical records for those sent a questionnaire it is impossible to know response rates in the different ethnic groups. However, it is known that respondents from output areas with increasing proportions of non-white people are less likely to respond. There remains the possibility that any differential response rates may introduce some bias which we are not able to allow for. If survey responders are more proficient in English, this may underestimate the communication difficulties experienced by certain minority ethnic groups.

Finally, as no objective measure of GP-patient communication exists for these data, our analyses are not able to provide insight into whether reported experience varies as a result of differences in actual experience or differences in reports of experience as a result of variations in expectations or survey response tendencies: for this, experimental approaches are required, as described in Chapter 6.

Implications for practice

The existence of marked differences in experience of GP-patient communication underlines the need for a renewed focus on those groups at risk of poorer quality of care. For practitioners, an awareness of the particular difficulties and frustrations encountered on both sides in cross-cultural consultations is an important first step. For patients with limited English language proficiency effective support for communication in the form of professional interpreters is important. [174] System-level as well as patient-targeted initiatives to improve health literacy are also likely to be important in reducing variations in care, although these inevitably require greater investment. [175]

Chapter 6. How do White British and Pakistani people rate communication within simulated GP-patient consultations? Experimental vignette study

Abstract

Background

Whilst minority ethnic groups have consistently reported poorer care in patient surveys, it is not known whether this is because they receive worse care or because they respond differently to such surveys.

Methods

We conducted an experimental vignette study to investigate whether South Asian people rate simulated GP consultations differently from White British people. 564 White British and 564 Pakistani adults were recruited using an in-home face-to-face approach. Trained fieldworkers completed Computer-Assisted Personal Interviews during which participants rated the communication within three video-recordings of simulated GP-patient consultations. Consultations were shown in a random order, selected from a pool of 16. Mean differences in communication score (on a scale of 0-100) between White British and Pakistani patients were estimated from linear regression.

Results

Pakistani participants, on average, scored consultations 9.8 points higher than White British participants (95% CI 8.0-11.7, p<0.001) when viewing the same consultations. When adjusted for age, gender, deprivation, self-rated health, and video, the difference increased to 11.0 points (95% CI 8.5-13.6, p<0.001). The largest differences were seen when participants were older (>55) and where communication was scripted to be poor.

Conclusions

Substantial differences in ratings were found, with Pakistani respondents giving higher scores to videos showing the same care. If we take these findings at face value, they would suggest that the lower scores reported by Pakistani patients in national surveys such as the GP Patient Survey represent genuinely worse care.

Introduction and rationale for the study

As outlined in Chapter 5, some minority ethnic groups have reported consistently lower patient experience scores compared to the majority population in both the UK and the US. [68, 140-143] Of particular concern within the UK, and confirmed by the analyses undertaken for this programme grant, South Asian groups report significantly more negative experiences of GP-patient communication than their White British counterparts. [120, 176] Potential explanations for these lower ratings focus on whether (a) South Asian patients receive lower quality care, or (b) whether they receive similar care, but rate this more negatively. [68]

A number of potential drivers of more negative ratings of similar standards of care exist. For example, it has been suggested that differences in the use of questionnaire response scales (e.g. [73]) may lead to South Asian groups being less likely to endorse the most positive options when asked to evaluate a doctor’s communication skills. Our analysis of GP Patient Survey data, drawing on item response theory to explore whether items receive systematically different responses from South Asian and White British groups, suggested that this was unlikely to be the case. [177] Yet there are also other, alternative drivers of poorer ratings of similar care, most notably that the evaluation of consultations by South Asian respondents is influenced by systematic variations in their expectations of, or preferences for, care.

Fundamentally, these concerns centre on a well-recognised and long-standing problem with surveys: that individuals may interpret and respond to the “same” question in many different ways. [178] Potential solutions to this problem arose first within the field of political science, where the use of standardised scenarios, or vignettes, was proposed to evaluate the disparity in responses to survey items. [75] Such approaches are particularly relevant to understanding minority ethnic experiences: as already described, alongside potential variations in scale use by individuals from various ethnic backgrounds, we also need to consider systematic cultural variations in expectations of or preferences for care, as well as the potential for systematic variations in actual experience. A recent US study adopted King’s vignette methodology to examine the extent of cross-cultural incomparability in survey responses, using predominantly written vignettes. [74] This online survey concluded that score variations observed on national surveys among African American, Latino and White respondents were likely to reflect true differences in real-life experiences, at least for items in the survey which used an “Always-to-Never” response scale. [74]

The aim of this strand of work was to build on previous vignette approaches to examine whether people from a Pakistani background rate the communication within simulated GP consultations differently from ratings provided by White British people. If these groups rate simulated consultations similarly when viewing identical video vignettes, then we would be able to conclude that it is more likely that the lower scores previously reported by South Asian respondents in national patient experience surveys reflect real differences in quality of communication within consultations.

Changes to study methods from original protocol

This strand of work, as stated in the original protocol, formed part of our wider aim of exploring in more detail the experiences of minority ethnic groups, together with the GP Patient Survey analyses reported above in Chapter 5:

“To understand the reasons why minority ethnic groups, especially South Asians, give lower scores on patient surveys compared to the White British population (aim 5).”

In our original protocol, to undertake this study we envisaged developing a DVD containing short clips (3-4 minutes long) of four simulated patient consultations, and asking respondents to rate these using the GP-patient communication items of the GP Patient Survey. These DVDs would be sent out, with questionnaires and instructions, to patients registered with practices with a high proportion of South Asian patients. We suggested using SANGRA (South Asian Names and Group Recognition Algorithm) to identify South Asian patients. In practice, we firstly devised a more robust and efficient approach to recruiting participants, using targeted face-to-face recruitment in partnership with the market research agency, Ipsos MORI. This enabled us to effectively reach a rigorously sampled set of participants of known Pakistani ethnicity. Secondly, participants rated simulated consultations during face-to-face computer-assisted interviews conducted by trained fieldworkers. This enabled us to collect high quality and consistent ratings of consultations. Our recruitment and rating approach is detailed in full below.

As we acknowledged in our original protocol, the requirement of the vignettes approach to show identical consultations to all participants meant that all videos had to be in English.

However, we had stated that, whilst we would therefore have to exclude patients who cannot understand English, we would make study questionnaires and documentation available in four Asian languages. As we employed face-to-face computer-assisted interviews in the study, this requirement was no longer necessary once we had screened for those who were confident in their ability in spoken English: this therefore represents a further improvement on our original study design.

Methods

In the experimental vignette study we undertook, we showed videos of simulated GP-patient consultations to White British and Pakistani respondents, who were asked to rate the quality of the communication within each consultation they viewed. The Improve study advisory group were particularly involved in consideration of the nature of the vignettes to be shown and the study materials.

Simulated consultations

To ensure generalisability and to avoid the chance inclusion of a characteristic or event which, unknown to us, might systematically be rated differently by the two participating ethnic groups, we produced a series of 16 vignettes. We set out to manipulate the vignettes on three key domains:

(1) the presenting complaint depicted within each consultation
(2) the quality of the communication within each consultation (poor or good), and
(3) the ethnic background of the actors playing the doctor and patient (South Asian or White British).

Published recommendations for the production of vignettes emphasise the importance of developing a valid script and considering how best to manipulate this on the domains of interest. [179] We therefore based our vignettes on real-life consultations video recorded as part of another workstream (the association between patients', raters' and GPs' assessment of communication in a consultation, for which we recorded over 500 real-life consultations). We undertook an extensive process of script development, roleplaying, and rating prior to filming the vignettes with professional actors (see Figure 14 and below for more details).

VIGNETTE DEVELOPMENT

RATIONALE

To vary vignettes on three domains:
1. The presenting complaint
2. The quality of GP-patient communication (poor or good)
3. The ethnic backgrounds of the doctor and patient (South Asian or White British)

CLINICAL CONTENT

- Derived from existing bank of 500+ videoed GP-patient consultations
- Identified consultations with unisex presenting complaints lasting less than 7.5 minutes (n=29)
- Four consultations selected by the research team: tennis elbow, persistent cough, numbness and perforated ear drum

SCRIPT DEVELOPMENT

- Each of the four clinical scenarios summarised for simulated patients on a proforma covering patient sociodemographics, clinical details, patients perspectives, past medical history and social history
- Each scenario roleplayed with a GP (JBe) and simulated patient in two versions – “good” and “poor” for communication. All roleplays video-recorded
- Communication quality of each roleplayed consultation scored using the GCRS by one rater (JS) to confirm the difference between “good” and “poor” communication roleplays at this stage
- Roleplayed scenarios transcribed in full to act as scripts for the vignettes. Minor changes to the content and stage directions were made by the research team in consultation with the simulated patient

VIGNETTE FILMING

- Vignette actors were recruited via an acting agency specialising in simulated patient roleplay
- Briefing packs were prepared for all actors, to include scripts, background summaries of vignettes, and verbal and non-verbal behaviour guides
- Vignettes were filmed with a professional film crew over two days, involving ten actors and one acting supervisor. At least one GP was present at all times to ensure clinical consistency
- Following filming, vignettes were professionally edited to create the final 16 films

FINAL VIGNETTES

- 16 films ranging from 2 to 8 minutes
- 8 films involving a South Asian GP-patient actor pairing: 4 different clinical scenarios, each filmed in two versions, poor communication and good communication
- 8 films involving a White British GP-patient actor pairing: 4 different clinical scenarios, each filmed in two versions, poor communication and good communication

RATING VIGNETTES

- Each vignette rated by three trained GCRS raters (all GPs) to determine its score for the quality of communication in relation to professionally-defined norms

Figure 14. Development of vignettes

The vignettes we produced covered four different clinical scenarios: persistent cough, perforated ear drum, painful elbow, and generalised numbness. We developed two different scripts for each clinical scenario: one designed to illustrate poor communication by the doctor, and one designed to illustrate good communication. We formulated “poor” and “good” standards of communication according to the Global Consultation Rating Scale (GCRS). [121] This observer-rated measure of communication competence (derived from the widely used Calgary-Cambridge guide to the medical interview [122, 123]) was developed as part of our workstream on patients’ and raters’ assessments of communication competence within a consultation. The GCRS instrument covers 12 domains including ‘initiating the session’, ‘gathering information’, ‘building the relationship’ and ‘achieving a shared understanding’ (see Appendix 12 for the full instrument). We then used both the ‘poor’ and ‘good’ versions of the four clinical scenarios to film two sets of vignettes. The first set of vignettes had White British actors playing the GP and patient, whilst the second repeated the same scripts but with South Asian actors playing the GP and patient. The GP role was acted throughout by either one White British or one South Asian actor; eight different actors (four White British and four South Asian) role-played patients, each participating in one clinical scenario. The final sixteen videos were each scored by three trained clinical raters using the GCRS to assess communication quality in relation to professionally-defined norms. [121] Mean GCRS scores for the “poor” communication vignettes ranged from 0.6 and 2.4 (out of 10), whilst mean GCRS scores for the “good” communication vignettes mean scores ranged from 5.1 to 8.4.

Data collection

Ipsos MORI fieldworkers conducted data collection, in collaboration with our team. As per the original protocol, we aimed to recruit 1,120 respondents, each of whom was asked to rate three simulated GP-patient consultations. Our original sample size calculation was based on data from the General Practice Assessment Questionnaire (which includes some identical items to the GP Patient Survey); we repeated this using more recent GP Patient Survey data. This confirmed that the inclusion of 560 Pakistani respondents and 560 White British respondents would give over 80% power to detect a 3.1 point difference (on a 0-100 scale) seen between these two groups after controlling for age, gender, deprivation, self-rated health and practice. As our analyses of GP Patient Survey data had identified that ethnic disparities were largest in older ages, we set out to recruit equal numbers above and below the age of 55 within each ethnic group.[176]

Following consultation with Ipsos MORI, we used different recruitment strategies for the different ethnic groups. To recruit Pakistani respondents, we selected output areas (geographically confined areas of approximately 130 households) in which at least 35% of the populations were identified as Pakistani in 2011 Census data. These were then ranked according to the proportion of the population aged over 50 (the cut-point of 50 years used for sampling reflects available census categories: for our recruitment, we specifically used a cut-point of 55 years and over). Trained fieldworkers then recruited participants within these areas using an in-home face-to-face approach, starting in the output areas with the highest proportion of residents over the age of 50. Fieldworkers were also provided with one or two output areas neighbouring to the area sampled, and were able to recruit from these if necessary. Snowball recruitment (for example, known neighbours suggested to fieldworkers) and additional household interviews were allowed.

To recruit White British participants, we first excluded output areas with low proportions of White British residents (less than 90%) and residents aged over 50. The remaining output areas were ranked by social grade (the percentage of people who were Social Grade A/B according to 2011 Census data) and geography. Ipsos MORI then selected output areas to approach using proportional systematic sampling.

Fieldworkers screened potential participants for ethnicity (using the Office for National Statistics 18-group categorisation) and for English-language competency (using a screening question regarding self-reported confidence in understanding short videos in English). Eligible respondents who consented then completed a Computer-Assisted Personal Interview (CAPI) during which the fieldworker used a standardised script. Each participant viewed three of the sixteen simulated consultation videos we had produced. Following each video, the participant was asked to rate the consultation using five GP-patient communication items taken from the most recent national GP-Patient Survey (Box 6). We assigned videos so that each participant saw three different presenting conditions, with at least one of the videos featuring each of the two ethnic GP/patient pairings, and at least one of the videos scripted to feature each of the two levels of GP-patient communication. The selection of videos shown to each participant was such that approximately equal numbers of all possible combinations were used, given the restrictions we have described. Participants also completed basic socio-demographic questions (age, self-rated health, whether born in the UK, language spoken most often at home). An area-based measure of socio-economic deprivation (Index of Multiple Deprivation) was recorded based on the participants' postcode.

Thinking about the doctor you have just seen in the video, how good was the doctor at:

	Very good		Good		Neither good nor poor		Poor		Very poor		Doesn't apply *
Giving enough time	☐	...	☐	...	☐	...	☐	...	☐	...	☐
Listening to	☐	...	☐	...	☐	...	☐	...	☐	...	☐
Explaining tests and treatments	☐	...	☐	...	☐	...	☐	...	☐	...	☐
Involving in decisions about your care	☐	...	☐	...	☐	...	☐	...	☐	...	☐
Treating with care and concern	☐	...	☐	...	☐	...	☐	...	☐	...	☐

* Considered to be uninformative for the purposes of our analysis

Box 6. GP communication items used to rate vignettes

Analysis

As in our previous analyses of GP Patient Survey data, we scored each participant's rating of each consultation by linearly scaling the response options between 0 (very poor) and 100 (very good) and averaging all informative answers when at least three of the five items were completed. We used linear regression to model the mean difference between White British and Pakistani participants' ratings of doctor-patient communication. We estimated the unadjusted difference in ratings, as well as the difference adjusting for patient age, gender, self-rated health, deprivation, and a set of 15 indicator variables for the video. We did not originally plan to conduct any analysis of interaction terms. However, the effect size found was much larger than that anticipated in our original power calculations, and so we investigated interactions between participant ethnicity and the following variables:

(a) relating to the video: ethnicity of GP/patient and quality of GP-patient communication
(b) relating to the participant: age, gender, and deprivation.

When modelling interactions, we used only variables for the video attributes, rather than using indicator variables for all videos. For interactions involving age, the oldest two age groups were combined and a continuous version of age-groups was used in the interaction term only. Confidence intervals and p-values were estimated using bootstrapping with 500 replications (given non-normal data), clustered by participant (with each participant supplying three communication scores). We conducted a sensitivity analysis that clustered the bootstrap resampling by output area rather than by participant to account for multiple sampling in households and small geographic areas: however, this made only trivial changes to standard errors and we consequently do not report this here.

Results

Participants

We recruited a total of 1128 participants: 564 (50%) self-identified as White British and 564 (50%) self-identified as Pakistani. The sociodemographic profile of participants is shown in Table 19. While the sampling restriction that half of participants in each group be 55 or above increased the similarity of the groups' age distribution, Pakistani participants were younger then the White British participants within the sampled age strata. Pakistani participants were also more likely to be male (58% vs 52%); to be in fair or poor health (38% vs 26%); and to live in the most deprived areas (82% vs 14%). Figure 15 shows the geographic locations where participants were recruited. White British participants were recruited from a wide range of geographic locations, whilst, as a result of our sampling approach, Pakistani participants were located from a small number of geographically confined locations. Between 202 and 220 participants scored each of the video vignettes for GP-patient communication (Table 20).

Table 19. Socio-demographic profile of vignette study participants

		All		White British		Pakistani	
		n	%	n	%	n	%
Age	18 to 24	88	7.8%	40	7.1%	48	8.5%
	25 to 34	154	13.7%	56	9.9%	98	17.4%
	35 to 44	151	13.4%	70	12.4%	81	14.4%
	45 to 54	175	15.5%	118	20.9%	57	10.1%
	55 to 64	267	23.7%	94	16.7%	173	30.7%
	65 to 74	179	15.9%	109	19.3%	70	12.4%
	75 to 84	95	8.4%	63	11.2%	32	5.7%
	85 or over	19	1.7%	14	2.5%	5	0.9%
Gender	Male	583	51.7%	255	45.2%	328	58.2%
	Female	545	48.3%	309	54.8%	236	41.8%
Self-rated health	Excellent	132	11.7%	82	14.5%	50	8.9%
	Very good	289	25.6%	181	32.1%	108	19.1%
	Good	348	30.9%	157	27.8%	191	33.9%
	Fair	207	18.4%	86	15.2%	121	21.5%
	Poor	152	13.5%	58	10.3%	94	16.7%
Deprivation	1 – Least deprived	108	9.6%	100	17.7%	8	1.4%
	2	137	12.1%	137	24.3%	0	0.0%
	3	122	10.8%	111	19.7%	11	2.0%
	4	221	19.6%	138	24.5%	83	14.7%
	5 – Most deprived	540	47.9%	78	13.8%	462	81.9%


White British

Pakistani

Figure 15. Geographic location of the census based output areas where participants were recruited

Table 20. Description of all video vignettes used with the number of times each video was scored

Video number	Clinical scenario	Scripted communication quality	Ethnicity of Dr and patient	Number of times video scored
1	Persistent cough	Bad	White	220
2			Asian	202
3		Good	White	202
4			Asian	212
5	Perforated ear drum	Bad	White	210
6			Asian	206
7		Good	White	217
8			Asian	207
9	Painful elbow	Bad	White	206
10			Asian	210
11		Good	White	210
12			Asian	215
13	Generalised numbness	Bad	White	216
14			Asian	222
15		Good	White	212
16			Asian	214

Main results

The distribution of communication scores for White British and Pakistani participants was skewed in both groups: however, the communication scores from Pakistani participants were typically higher than those from White British participants (Figure 16). The mean communication score from Pakistani participants was 67.3 out of 100, 9.9 points higher (95% CI 8.0, 11.7, $p<0.001$) than the mean score from White British participants (57.4 out of 100).

In a regression model (full output shown in Table 21) adjusting for participant age, gender, self-rated health, deprivation, and video, there was a slightly larger difference between the two ethnicities: 11.0 points (95% CI 8.5, 13.5, p<0.001).

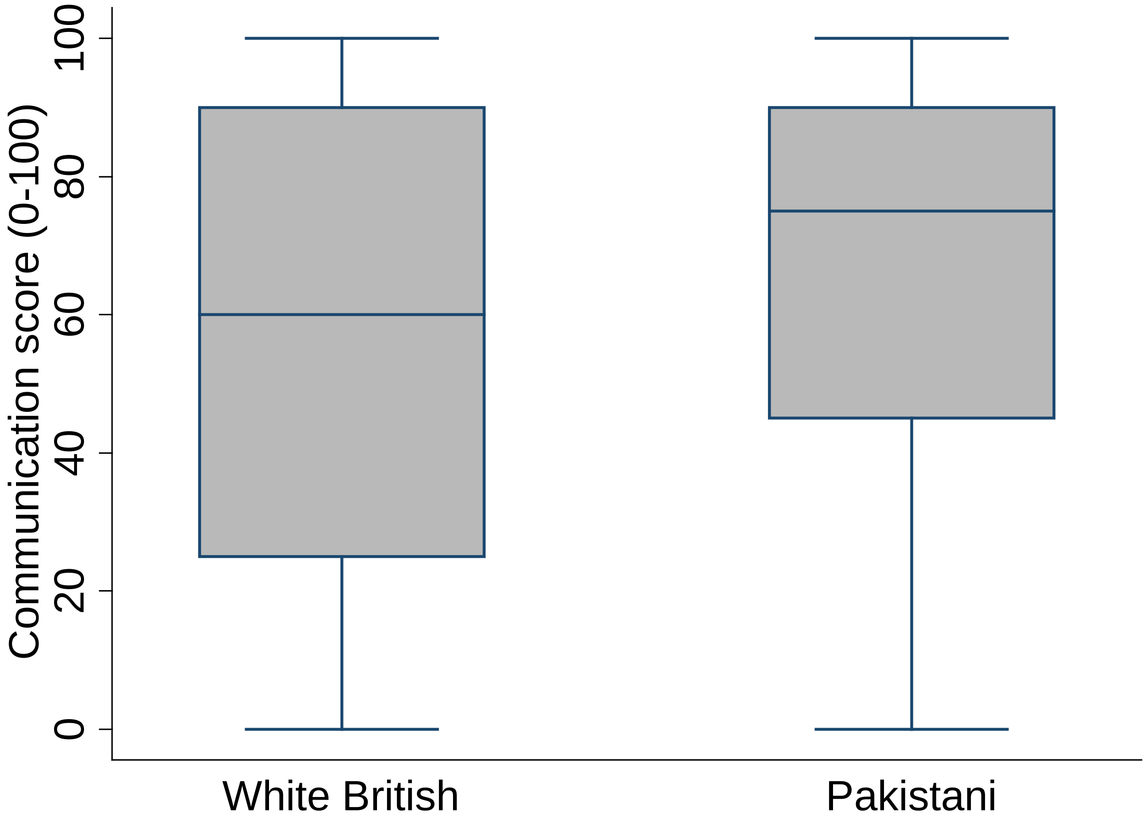


Figure 16. Box plots showing the distribution of GP communication scores recorded by White British and Pakistani participants

Table 21. Output from the main regression model adjusting for socio-demographic characteristics but with no interactions

		Adjusted difference (95% CI)	P-value
Ethnicity	White British	Reference	<0.001
	Pakistani	11.01 (8.53, 13.49)	
Age (Years)	18 to 24	-5.55 (-8.94, -2.16)	<0.001
	25 to 34	-4.96 (-7.99, -1.93)	
	35 to 44	-1.67 (-4.67, 1.33)	
	45 to 54	-1.86 (-4.60, 0.87)	
	55 to 64	Reference	
	65 to 74	4.01 (1.20, 6.83)	
	75 to 84	6.70 (3.26, 10.13)	
	85 or over	3.66 (-3.66, 10.97)	
Gender	Male	Reference	0.115
	Female	1.41 (-0.34, 3.16)	
Self-rated health	Excellent	Reference	0.866
	Very good	-1.15 (-4.05, 1.74)	
	Good	-1.65 (-4.71, 1.41)	
	Fair	-1.77 (-5.12, 1.58)	
	Poor	-1.41 (-5.21, 2.38)	
Deprivation	Least deprived	Reference	0.505
	2	-0.92 (-4.10, 2.27)	
	3	1.08 (-2.26, 4.42)	
	4	-1.45 (-4.57, 1.68)	
	Most deprived	0.13 (-3.32, 3.58)	
Video number	1	Reference	<0.001
	2	-3.90 (-6.79, -1.01)	
	3	-56.51 (-60.51, -52.50)	
	4	-49.57 (-53.77, -45.37)	
	5	-4.09 (-7.06, -1.12)	
	6	-7.45 (-10.58, -4.33)	
	7	-48.08 (-51.81, -44.34)	
	8	-49.70 (-53.53, -45.87)	
	9	-3.24 (-6.33, -0.14)	
	10	-7.40 (-10.48, -4.33)	
	11	-52.19 (-56.03, -48.34)	
	12	-48.94 (-52.80, -45.08)	
	13	-9.59 (-12.89, -6.29)	
	14	-9.36 (-12.45, -6.27)	
	15	-54.23 (-58.07, -50.38)	
	16	-46.63 (-50.52, -42.73)	

Analysis of interactions

As the difference in scores between Pakistani and White British participants was considerably larger than that expected at the design stage, we were able to explore interactions between ethnicity and other variables. We found no evidence that the difference in scores between Pakistani and White British participants varied by patient gender (p=0.92), deprivation (p=0.68), or by the ethnicity of the doctor/patient actor pairing in the videos (p=0.53). There was, however, strong evidence that the difference in scores between Pakistani and White British participants was larger for older participants (p=0.001), and for consultations scripted to contain poorer doctor-patient communication (p<0.001). Table 22 shows the mean difference in age by good/poor scripted communication strata, estimated from a model containing all main effects, plus (1) ethnicity and age interactions, (2) ethnicity and good/poor communication interactions, and (3) the three-way interaction between those variables (p<0.001 for three-way interaction).

Table 22. Adjusted difference in communication scores for age group by good/poor scripted communication between White British and Pakistani participants. A positive difference implies Pakistani patients gave, on average, higher (more favourable) scores.

Age	**Scripted communication**	
	Good	**Poor**
18 to 24	-1.31 (-5.38, 2.76)	10.29 (5.00, 15.57)
25 to 34	-0.15 (-3.58, 3.27)	13.32 (9.10, 17.54)
35 to 44	1.01 (-1.96, 3.97)	16.34 (12.91, 19.77)
45 to 54	2.17 (-0.62, 4.95)	19.37 (16.24, 22.50)
55 to 64	3.33 (0.39, 6.27)	22.40 (18.94, 25.86)
65 to 74	4.49 (1.11, 7.87)	25.42 (21.16, 29.69)
75 and over	5.65 (1.64, 9.66)	28.45 (23.11, 33.79)

The difference between scores given by younger (under 55 years) White British and Pakistani participants to consultations containing “good” communication was small and not statistically significant. However, larger and statistically significant differences were seen for older patients and for consultations portraying “poor” communication at all ages. In these “poor” consultations, the difference in scores increased with rising age of participants. For example, ratings of consultations with poor communication were 10.29 points higher (95% CI 5.00, 15.57) for 18 to 24 year old Pakistani participants than White British participants of the same age. This difference increased to 28.45 points (95% CI 23.11, 33.79) for over 75 year olds.

Discussion

This experimental study found that respondents from a Pakistani background rated simulated GP consultations substantially more positively than their White British counterparts. These differences were largest for consultations depicting poor doctor-patient communication, and for older respondents. The differences we observed were in the opposite direction to those in the national GP Patient Survey, which relates to a patients’ most recent consultation with a GP; for which Pakistani respondents give significantly lower scores for communication than their White British counterparts.

Strengths and limitations

We used an in-home face-to-face recruitment approach to ensure access to a wide range of respondents, independent of the GP practice they were registered with. However, it is possible that respondents who agreed to participate in this research may differ in a number of unidentified ways from the population as a whole. For example, to ensure efficient recruitment to the study, we focussed our efforts on high-density Pakistani areas, which also have high levels of deprivation (the 82% of participants living in areas in the most deprived quintile compares to 51% nationally). The sampled population may, therefore, differ from the Pakistani population as a whole: for example, recent research suggests that minority ethnic populations in lower ethnic density areas may report higher satisfaction with health care.[180] Ratings of consultations by “analogue patients” (members of the public asked to rate care received by a third party), such as our participants, are commonly more critical than patients commenting on their own care. [181] In our study, negative response options were used more often than in the national GP Patient Survey: for example, only 2.6% of answers to the GP communication questions in the most recent GP Patient Survey were given as poor or very poor, compared to 26.6% of answers in this study. [182] We deliberately set out to create a wider-than-typical range of communication quality within our vignettes to enable us to test the hypothesis that differential response tendencies between groups may only exist at one end of the communication range – for example, that South Asian respondents tended to be more negative about the best care but rate the poorest care in the same way. In fact, we found that there were differences in ratings (with Pakistani respondents more positive) at both ends of the communication spectrum, reinforcing our interpretation that the disparities in real-life surveys are not to do with differential response tendencies .To enable the same vignettes to be viewed by all participants, the study was conducted in English, limiting our ability to understand evaluations by those with low English language proficiency (and who might, for example, respond to the GP Patient Survey in other languages). In the US, minority ethnic groups preferring languages other than English generally show response tendencies that are in the same direction as English-preferring members of the same minority ethnic, but to a greater extent, perhaps reflecting a continuum of acculturation. [140] However, it was not possible to produce vignettes that would remain equivalent in other languages, and as 99.8% of respondents to the GP Patient Survey respond in English, our ability to extrapolate to the wider population remains high.

Previous examinations of inequalities in patient experience between ethnic groups (including our own) have commonly relied on real-world data, in which it is difficult to distinguish whether differences are attributable to variations in care or variations in the reporting of that care. [68, 120, 140-143, 176] The experimental design we used in this workstream enabled us to control the content of the consultations being rated by respondents in order to explore how differences in reporting may explain the disparities in minority ethnic experience in real-life surveys. It builds on previous vignette research by using multiple video vignettes manipulating several key attributes. [74, 75] Video vignettes have so far been little employed in this field, in spite of evidence of viewers perceiving them as realistic and enabling immersion in the situation at hand. [179] In the US, Weinick et al. reported no evidence of differences among White, African American and Latino evaluations of doctor-patient communication in vignettes when using an “Always-to-Never” response scale; they concluded that variations within national surveys on such items for these groups were likely to reflect differences in real-life experiences. [74] In this study, however, we found substantially more positive ratings by Pakistani in comparison to White British respondents.

Implications for practice

We designed this workstream to explore whether people from a Pakistani background rate the communication within simulated GP consultations differently from ratings provided by White British people. Similar ratings of simulated consultations from both ethnic groups would have suggested that the low scores observed in national surveys from Pakistani and other South Asian respondents reflect real differences in the quality of communication experienced by these patients in comparison to White British patients. The substantially more positive ratings from Pakistani respondents that we observed in our experimental study suggest that not only are there differences in the quality of communication in real-life consultations, but that these differences are even greater than those identified in real-life surveys. We suggest that Pakistani patients receive genuinely worse standards of communication within a consultation. However, whilst we can be confident that differences in experience exist, it is difficult to extrapolate our vignette-derived data to estimate the magnitude of difference in real life. Poor communication for these groups may arise from system-level, provider-level and/or patient-level factors. [183] For example, language barriers within consultations may lead to more negative experiences of care for both doctors and patients. [164] Levels of acculturation may be linked with a patient's ability to navigate the health care system, with consequent impacts on patient experiences of care. [166] System- and provider-level issues, including discrimination and bias, are sensitive and challenging topics, but ones to which more recent dialogue has looked as likely key contributors to inequalities in care. [171]

Our findings add substantial weight to the likelihood that inequalities affecting South Asian people in national surveys reflect systematic variations in the quality of communication within consultations. Whilst there is a body of research into the drivers of inequalities in care, we suggested that further research in this area now needs to focus on how factors including language barriers, health literacy, provider-side discrimination and system-level failures combine to inhibit good communication within individual consultations.

SECTION C

Using data on patient experience for quality improvement

Chapter 7. Attitudes to receiving feedback from patient experience surveys: focus groups with practice staff

This material is reproduced, from Boiko, O., Campbell, J. L., Elmore, N., Davey, A. F., Roland, M., & Burt, J. The role of patient experience surveys in quality assurance and improvement: a focus group study in English general practice. Health Expectations, 2015, 18(6), 1982-1994,[277] *under the terms of the Creative Commons Attribution Non-Commercial No Derivatives License CC BY-NC-ND and permits non-commercial use, distribution and reproduction in any medium, without alteration,*

Abstract

Background

Despite widespread adoption of patient feedback surveys in the NHS, evidence of a demonstrable impact of surveys on service improvement is sparse. The objective of this study was to explore the views of primary care practice staff regarding the utility of patient experience surveys.

Methods

We conducted focus groups with staff from 14 practices following the receipt of feedback from a recent patient experience surveys.

Results

Whilst participants engaged with feedback from patient experience surveys, they routinely questioned its validity and reliability. Participants identified surveys as having a number of useful functions: for patients, as a potentially therapeutic way of getting their voice heard; for practice staff, as a way of identifying areas of improvement; and for GPs, as a source of evidence for professional development and appraisal. Areas of potential change stimulated by survey feedback included redesigning front-line services, managing patient expectations, and managing the performance of GPs. Despite this, practice staff struggled to identify and action changes based on survey feedback alone.

Conclusions

Whilst surveys may be used to endorse existing high quality service delivery, their use in informing changes in service delivery is more challenging for practice staff. Drawing on the Utility Index framework, we identified concerns relating to reliability and validity, cost and feasibility, acceptability and educational impact which combine to limit the utility of patient survey feedback. Feedback from patient experience surveys has great potential. However,

without a specific and renewed focus on how to translate feedback into action, this potential will remain incompletely realised.

Introduction and rationale

As outlined in the introduction to this report (chapter 1), feedback from patients is intended to inform quality improvements by increasing the responsiveness of the health care system to the needs of service users, and by identifying areas of poor performance or organisation which might be susceptible to change. [78, 79, 129] Whilst policy initiatives such as the introduction of the Quality and Outcomes Framework or revalidation highlight feedback on patient experience as a key driver of quality improvement, evidence suggests patient experience has had only limited impact on service delivery, [42] and GPs and other healthcare professionals may experience difficulties in making sense of survey-generated information.[111, 184]

In this strand of work, we draw on qualitative data to examine how primary healthcare practitioners and their teams view and act upon feedback from patient experience surveys. We examine the role that patient feedback plays in both assessing and improving standards of care. To assist our consideration, we adopted van der Vleuten's Utility Index model as the basis for considering potential drivers of the gap between receiving and acting on patient feedback in primary care practices. [185] Originally developed as a framework for assessment design and evaluation in educational settings, reports of the use of the utility model have been extensive, although such reports have nearly always emanated from educational settings. We felt that the six domains of the model [186] (educational impact, validity, reliability, cost, acceptability and feasibility) also had potential relevance when considering issues relating to the introduction and use of surveys of patients experience of care in routine clinical settings.

Changes to study methods from original protocol

The aim of this strand of work, as stated in the original protocol, was:

"To understand how general practices respond to low patient survey scores, testing a range of approaches that could be used to improve patients' experience of care (aim 1)."

In our application, we envisaged drawing on GP Patient Survey scores to facilitate a discussion with participating practices on their most recent results, their responses to their scores, and any intention to change as a result. However, as part of our overall programme of work we conducted an individual GP-level postal survey with 25 practices (aim 2: see

Chapter 9). This gave us the opportunity to feedback results to practices and individual GPs from our own, very recent survey, and subsequently explore their responses to both practice-level feedback and the potential for individual feedback within focus groups (reported in this Chapter) and interviews (reported in Chapter 8).

Methods

We conducted a postal survey of patients who had recently seen a doctor at one of a stratified random sample of 25 practices in Cornwall, Devon, Bristol, Bedfordshire, Cambridgeshire and North London (see Chapter 9 for details of sampling, recruitment and survey conduct). The patient experience survey used was based on items from the national GP Patient Survey, and asked patients about access, waiting times, opening hours, and continuity and interpersonal aspects of care. [187] We reported results back to practice staff at aggregate practice level (report to all staff) and at individual GP level (confidential reports to each participating GP).

We purposively approached practices who had participated in the survey to take part in focus groups, aiming to reflect a diversity of practice size, geographical location and practice-level survey scores for communication. We undertook focus groups in fourteen practices. Groups (with between four and fifteen participants in each) were conducted following the completion of practice surveys and feedback of the findings to staff. Overall, 128 professionals from a range of backgrounds (40 GPs, 18 managers, 18 nurses, 20 receptionists, 13 administrators and secretaries, and 19 other staff including dispensers and health care assistant) took part. In reporting, all practices were assigned a practice pseudonym: real practice names were not used (Table 23).

Table 23. Participating practices and focus group participants

Practice pseudonym	2009/10 national GP Patient Survey scores for communication	Location	No. of practicing GPs	No. of focus group participants
Highfields	High	rural	4	5
Church Road	High	urban	8	15
Fieldview	High	rural	5	9
Town Road	Medium	city	3	11
Meadow	Medium	rural	5	13
Pilkington	Medium	urban	3	9
The Towers	Low	urban	2	4
Brentwell	Low	city	5	4
Crossways	Low	city	7	6
White Road	Low	urban	2	7
Torch Street	Low	city	6	10
The Maples	Low	urban	5	13
Fallowfield	Low	city	4	6
Beeches	Low	urban	5	15

Reproduced, from Boiko, O., Campbell, J. L., Elmore, N., Davey, A. F., Roland, M., & Burt, J. (2015)[277]

Focus groups, lasting around an hour, were held on practice premises and were facilitated by experienced qualitative researchers. A second researcher was present at each group to take notes. We piloted a topic guide (Box 7) at two non-study practices prior to beginning fieldwork. Key areas of discussion included attitudes to patient surveys, past experiences of surveys, and practice procedures for dealing with survey feedback. All groups were transcribed verbatim. To maintain anonymity, participants were assigned pseudonyms.

Items from the focus group topic guide

- What do you think of patient surveys in general? What do you think the survey results are saying to your practice?
- Are the results of patient surveys circulated within your practice and if so, to whom? Have the scores encouraged you or your colleagues in wanting to change anything?
- Do you think that individual GP scores following a patient experience survey could have an impact on the practice as a whole?
- Do you think that over time, surveys of patient experience which focus on individual doctors' skills, might affect the attitude of doctors towards their patients – or the attitude of patients towards their doctors?
- To further explore the impact of individual GP performances on practice functioning, focus group participants were also invited to comment on two hypothetical situations where some doctors within the practice received less favourable scores from patient surveys than other doctors.

Box 7. Sample focus group questions

Reproduced, from Boiko, O., Campbell, J. L., Elmore, N., Davey, A. F., Roland, M., & Burt, J. (2015)[277]

We drew upon framework approaches to organise and analyse our data, which allowed for themes to be assigned both from *a priori* research questions and from the narratives of focus group participants. NVivo v.10 software (QSR International Ltd, 2012) was used for organising and examining data. Analysis was undertaken by two researchers (OB and JB) and took place over five stages: familiarisation (reading transcripts and listening to recordings in detail to gain an overview of content), thematic analysis (developing a coding scheme), indexing (applying the codes systematically to the data), charting (re-arranging the data according to the thematic content to allow comparative analysis), and mapping and interpretation (defining key concepts, delineating the range and nature of phenomena, creating typologies, findings associations, providing explanations and developing strategies). 188

Guided by this approach, we drew on transcripts from the first focus groups to develop an initial coding framework. This included 48 codes grouped loosely into headings including validity of surveys, interpretation of survey feedback, organisational changes and performance comparison. Our coding framework went through a process of application, discussion and revision until all transcripts were coded using the final agreed version. Codes were subsequently grouped into four overarching analytical themes: survey validity and interpretation, practice dynamics, leadership and interprofessional decision-making, and improvement strategies. The coding of each theme and subtheme were further triangulated by two researchers against a number of transcripts and discussed within the wider research team. The study was guided by our advisory group including four patient and public involvement members, who provided input into study design and conduct and interpretation of findings.

Findings

We focus in our findings on the organisational response of practice staff toward patient surveys. First, we consider how practice staff understand and engage with surveys and survey feedback. Then, we consider three dimensions of potential and actual change which appear to have been driven, in full or part, by surveys: redesigning front-line services, managing patient expectations, and managing the performance of GPs. In the discussion, we place our findings within the context of the Utility Index model to consider how the utility of surveys to practice staff might influence their uptake as either quality assurance or quality improvement mechanisms. [185]

1. Understanding of, and engagement with, surveys

All practice teams had extensive, first-hand involvement in surveying their patients, and in receiving feedback from the GP Patient Survey. Attitudes to patient surveys were contradictory. Recent experiences of payments linked to survey results under the Quality and Outcomes Framework had caused resentment for many, particularly those who had lost out financially. Overall, practice staff found it difficult to trust surveys to reflect “reality”. Despite this, their expressed ambivalence about surveys was often mixed with an interest in and engagement with the findings. We explore these ideas in more detail below.

Credibility of surveys

Practice teams spoke broadly about the perceived weaknesses of survey methods, singling out issues around their design, administration, representativeness, reliability, sample size, bias and the political ends which they were intended to serve:

> *"The surveys only take a snapshot."*
>
> *Nurse, Torch Street*

> *"Only people with strong views complete them."*
>
> *Receptionist, Crossways*

> *"You need to have sufficient sample size and a meaningful way of comparing across different GPs in order for someone to get some useful knowledge out of it."*
>
> *GP, Fallowfield*

Practice staff sometimes struggled with the concept of quantifying patient experience, suggesting that the complex reality of healthcare interactions could not be measured using such rigid methods:

> *"And a lot of this data that's collected in a measurable kind of way doesn't really represent reality. There's kind of a fixation on measurable outcomes, but they don't really tell us what's going on, they're just measuring that thing."*
>
> *GP, The Maples*

Discussions often distinguished between the utility and relevance of different types of surveys, from in-house surveys conducted by receptionists handing out questionnaires, to the national survey programme. Local surveys were seen as enabling practice staff and patients to have greater control over the perceived relevance of the questions, although teams were often cynical about their robustness:

> *"And some practices can manipulate their patients that they survey, so they will only hand out the questionnaire to nice patients and patients they know, they won't do it on duty day when doctor is maybe running behind or very busy."*
>
> *GP, Church Road*

Criticisms levelled at the current national GP Patient Survey included its distribution to a sample of all patients registered with a practice regardless of whether they have consulted recently, the focus on feedback at practice rather than individual practitioner level, and the absence of free text comments. Surveys that encompassed these elements were frequently regarded more positively:

> *"We want to see data tailored to individual practitioner, because we all practice differently".*
>
> *GP, Town Road*

Other sources of patient feedback, such as complaints, were often framed as a more useful source of information to understand where the problems lie:

> *"And I think we learn a lot more from patients that write to us individually with complaints."*
>
> *Administrator, Town Road*

Engaging with surveys

Despite these concerns, the importance attached to patient feedback via surveys was well recognised and broadly accepted:

> *"I think we must not be too negative about surveys because they are part of the way we do things nowadays […] I think if you look at how general practice changed particularly over the last 20 years, it has become a lot more patient focussed and those things did not happen by accident, they have happened by design, and patient surveys have been a tool to drive that."*
>
> *GP, Highfields*

However, whilst participants (in particular GPs and practice managers) positively engaged with survey findings from year to year, contradictions and tensions were evident, for example in relation to the validity of patient's reports:

> *"I think it is the only way to find out exactly what's going on is to do a survey. The only way you really find out what the patients think. They are not always honest. Well, they are not always honest on the survey either."*
>
> *(Nurse, Beeches)*

> *"I think it is useful for the extremes, but personally, I don't think it is particularly useful for any middle ground. [Later in focus group] I think it's very useful, when it compares against national average. I find that really, really helpful."*
>
> *(GP, Beeches)*

For practices that scored below national benchmarks, engaging with survey findings was often an emotional experience:

> *"It can be a bit disheartening at times though, if you feel that you're really doing your best and then you get negative feedback."*
>
> *(Receptionist, Torch Street)*

The functions of surveys

In general, practice staff valued feedback from surveys as a source of information about their performance. Participants suggested that patients, individual GPs, and the practice as a whole might all benefit from surveys: for patients, for example, there may be a therapeutic function, *"the chance to get something off their chest and ... to then move on."* (GP, Highfields). For GPs, the function of surveys was often to fulfil the requirements for appraisal. For practice staff, surveys could have a clear 'improvement' message, including the potential to highlight under-performing GPs:

> *“It helps to highlight areas of improvement, to make sure that we’re continuing to do as well as we think we’re doing and it prevents us becoming complacent and assuming that you’re doing well. I mean if we are doing well, then it confirms that we are doing well, if we’re not doing well then it identifies areas that hopefully we can change. But not always.”*
>
> *GP, Highfields*

> *“You can argue over the validity of surveys but if over three/four years someone is consistently scoring low in certain areas, you can start making assumptions about the doctor performing not very well in the practice.”*
>
> *(GP, Brentwell)*

2. Changes driven by survey feedback

The processing of survey feedback by practice staff was the essential first step in making any changes. These might encompass re-designing frontline services, managing patient expectations, and managing the performance of GPs. However, variation was evident in how transparent practice staff were in sharing survey information within the team, and in whether practice-level feedback was circulated between GP partners, to just a few practice decision-makers, or to all of the staff. In a small number of practices, results had been fed back promptly by staff to their patient participation groups. Inevitably, the level of transparency impacted on the understanding of and engagement with patient feedback by practice staff.

Redesigning front-line services

Practice staff often described changes they had made to front-line services and systems as a result of patient preferences, including modifications to their facilities, appointment systems, and to staffing issues such as staff training. For example, car parks had been extended, GP triage introduced, and new call management programmes installed. Staff in three practices clearly articulated the incorporation of suggestions from patient surveys into an annual action plan. However, in most practices few changes were attributable to survey feedback, though the survey sometimes provided a "nudge" to action in areas practice staff had been already been considering:

> *Nurse:* *"We did a change to open extended hours Thursdays, so that is a good thing – a benefit from last year's I think, or was it the year before?"*
>
> *Receptionist:* *"Yeah, a year now."*
>
> *GP:* *"Although it wasn't really a response to a survey, that, it was a response to an initiative from…It was a response to the fact that there was funding available from the PCT for extended hours."*

(Torch Street)

Managing patient expectations

For staff in some practices, survey feedback raised issues about how to communicate change to patients, how to shape expectations, and how to raise patient responsibility. Practice staff often felt they struggled to respond to patient demands and to increase understanding amongst their patients about practice systems:

> *Facilitator: "Was there anything in the feedback where you kind of, you thought maybe you wouldn't respond?"*
>
> *GP1: "Opening Sundays."*
>
> *[laughter]*
>
> *GP2: "I think another thing that was highlighted, for instance, is the question of marketing. I think we probably haven't, in spite of having additional extended hours on Saturdays, and I think that was, was one of the things we had a big conversation about the MORI survey. At that point, we were offering all sorts of extended hours, but patients didn't seem aware of it."*

(The Maples)

Practice staff often felt that a perceived lack of understanding of systems and services was evident in "demanding" patients, whatever effort was made. Furthermore, issues that suited one group of patients (music in the waiting room, telephone consultations) ran the risk of provoking dissatisfaction in others.

Practice staff felt that patients had a role to play in smooth and efficient functioning of primary care services. Staff spoke about increasing patient accountability and engaging patients in the feedback process through patient participation groups.

Managing the performance of GPs

Individual GP performance was regarded as an important factor in determining overall practice scores. Several managers in low scoring practices admitted that, it was difficult to tackle individual doctor's (poor) performance:

> *Manager:* *"If the survey results are between (the survey providers) and the doctor, and he knows that or she knows it, there's absolutely no reason for them to change their ways, is there? What is the motivation to change, what is the driver to change when they have been rude or pretty lazy? Nobody knows that, let's get on and continue as before. It is only when this information becomes available to, perhaps, the practice, that things could start to change. And when I say practice, who in that practice I don't know, it could be the executive partner. But I think somebody ought to know and somebody ought to discuss these issues."*
>
> *Nurse:* *"What's the point in doing the survey anyway? If nothing is going to happen, is no point in doing that if doctor..."*
>
> *Manager:* *"Nothing is going to change."*
>
> *Nurse:* *"…got the bad score and they keep it to themselves."*
>
> *(Brentwell)*

The idea of having an ‘outlier’ doctor, whether it was a high or a low performer, was familiar to practice staff. Both scenarios could have an effect on the running of the practice, for example when patients found it difficult to obtain an appointment with a particularly popular doctor. In addition, the complexity and inter-linking of factors influencing patients’ responses was highlighted - patients’ overall impression of the surgery and of the appointment system was perceived to influence their reports regarding consultations, and possibly the performance of the doctor too:

> *“Looking at the way people have access, the way the practice is organised, that they have access to facilities within the practice, the hours that the practice is open, the stage of the practice, the receptionist, how the admin is done, virtually how the sort of machinery of the practice works... I would not be surprised that where you had a poorly organised practice, poor machinery, if you like, you also had poor doctors, because I think doctors are influenced by the machinery in which they work, as well as influencing the machinery* themselves.”
>
> *(GP, Pilkington)*

The majority of teams stressed that they would support a doctor who consistently received negative patient feedback, although they did raise concerns about the difficulty of having an ’unmanageable’ GP in the practice. Suggested internal mechanisms of support ranged from support via a team member, role-plays and peer support sessions, to interventions by a partner and/or manager. Creating a supportive environment was described as an important enabler, although it was not always clear what the concept of ‘supportive environment’ actually meant for the participants. There were no doubts that doctors who were put “at the bottom of the pile” by survey results could perceive any intervention as threatening. In three low-scoring urban practices, staff were supportive of making the doctors’ scores publicly available, identifying a responsibility to maintain patient safety.

Barriers to improvement

Discussions on potential improvements most commonly focussed on changes to practice premises and on organisational aspects of the delivery of care. Even for such changes, which may have been at least in part precipitated by patient survey feedback, staff in most practices felt there was little long-term impact on patient opinion:

> *“We've done a number of things and the Mori poll results have been remarkably stubborn in terms of the change in perception by patients. That’s been quite slow.”*

(Manager, Beeches)

As one respondent highlighted, survey fatigue and the feasibility of being able to make relevant, meaningful changes was a persistent problem:

> *“The cynicism that [Dr Ahmed], has quite rightly identified as being the problem with the surveys, is the fact that we have been surveying, and patients have been surveyed, for several years, the questionnaires are inevitably similar, the responses are inevitably similar, but the consequences of the survey are depressingly zero. So there may be a request from patients, for example, that old chestnut, the Saturday morning surgery, but that has never been, and never will be, as far as I'm aware […] funded to take place. So, you then question the validity, the point of actually having the survey.”*

(GP, Church Road)

Staff highlighted a wide range of barriers to implementing changes which may have been requested by patients, most particularly expressing concerns around funding and staff capacity. A distinction was made between patient “needs” and patient “wants”, with identification of an ongoing struggle to meet unrealistic expectations:

> *“It is a bit like opening on Saturday issue. Would you like the surgery to be open on Saturday? Yeah. Would you like us to go 24 hours? Yeah. Are you going to pay more taxes to have it open on Saturday? No. Are you going to use appointments during the week when you are able to make it? Mmm, not sure. But if the question is would you like to have it open on Saturday? Yeah. Consumerist.”*

(GP, Church Road)

There was far less discussion and agreement on how to effect changes to interpersonal aspects of care, if survey feedback highlighted issues relating to a particular GP. Issues included confidentiality and the ‘unlikely’ situation of GP feedback being shared with other practice staff, and the idea that practice staff may need to recognise a balance in a GPs’ interpersonal abilities and other aspects of their professional practice (“maybe that doctor is not a great communicator but they are great at doing something else, you know” (GP, Church Road)).

Ultimately, staff in many practices felt there was little external support for making changes in response to patient feedback:

> *“… we need more support in this area […] one of my concerns up until now is that sometimes services have come out and there has been very little support from anyone to say, right this is how you can improve things that might help, or we understand why you might be having problems, which ways we can help you with that. It has always been: here is your survey results, it is up to you how you sort it.”*
>
> *(GP, Highfields)*

Discussion

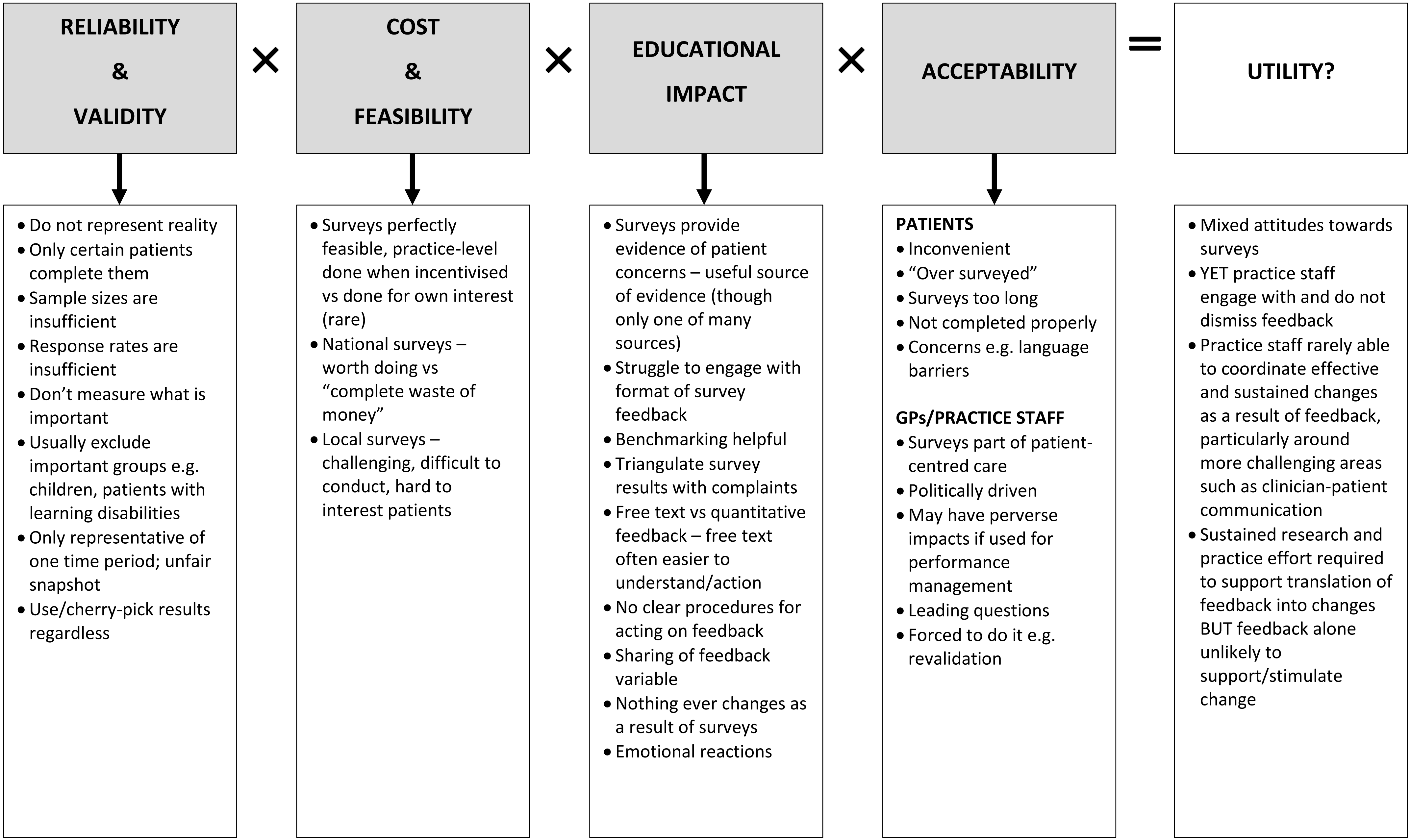
Our findings suggest two primary purposes of large scale surveys of patient experience, as identified by practice staff. First, surveys may be used to endorse and affirm good clinical practice or service organisation. Second, in line with the aspirations of policy makers, surveys may provide evidence to inform improvements in healthcare provision. [78] Our findings suggest that staff in general practice broadly view the role of patient feedback as one of quality assurance, providing evidence of whether they are offering an acceptable level of care to their patients. However, the role of surveys in quality improvement appeared less certain amongst participants. Whilst we identified *potential* dimensions of change (for example, front-line service improvements, management of patient expectations, and management of GPs’ performance) which could be informed by survey feedback, actual changes were usually confined to ‘easy targets’ for modification such as décor or playing music. Practice staff appeared to vacillate between questioning the credibility of survey findings and taking them at face value: as we observed, respondents could be critical of survey methods whilst being pleased their practice had “done well”. For those who had done less well, pathways to change were not often clear. These organisational responses to patient experience surveys were, inevitably, dominated by GPs and practice managers – receptionists and administrative staff were far less vocal. Whilst not reported within here, our analyses suggest important variations in the extent of the influence of practice managers, and the dynamics between practice managers and GPs, on how practice staff as a whole reflect and act upon patient feedback.

Strengths and limitations

We drew on a large sample of primary care practitioners providing care in a range of practice settings, and representing a range of primary healthcare professionals. Fourteen focus groups, of varying size, acted, we believe, as an effective means of capturing a range of participant views. However, it is possible that the focus group approach meant that a number of staff, particularly those in more junior positions, felt unable to give their true opinions on the organisation and delivery of care within the practice. Here, face-to-face individual interviews may have uncovered more sensitive information particularly, for example, around approaches to dealing with individual staff members who receive poor patient feedback.

The Utility Index

Van der Vleuten's Utility Index was originally developed to consider assessments within an educational context (for example, the provision of feedback on progress to medical trainees or the conduct of examinations for specialist training). [185] However, we believe this model also has potential value in exploring the utility of patient surveys in service contexts. Examining our emerging findings through the utility lens, which we undertook as a post-hoc exercise, suggested that the overall value of patient feedback from surveys (and thus its potential to drive significant quality improvements) is potentially undermined by a combination of variable attitudes to the credibility of the feedback, and to challenges for practice staff in identifying and bringing about meaningful changes (Figure 17).



RELIABILITY & VALIDITY × **COST & FEASIBILITY** × **EDUCATIONAL IMPACT** × **ACCEPTABILITY** = **UTILITY?**

RELIABILITY & VALIDITY

- Do not represent reality
- Only certain patients complete them
- Sample sizes are insufficient
- Response rates are insufficient
- Don't measure what is important
- Usually exclude important groups e.g. children, patients with learning disabilities
- Only representative of one time period; unfair snapshot
- Use/cherry-pick results regardless

COST & FEASIBILITY

- Surveys perfectly feasible, practice-level done when incentivised vs done for own interest (rare)
- National surveys – worth doing vs "complete waste of money"
- Local surveys – challenging, difficult to conduct, hard to interest patients

EDUCATIONAL IMPACT

- Surveys provide evidence of patient concerns – useful source of evidence (though only one of many sources)
- Struggle to engage with format of survey feedback
- Benchmarking helpful
- Triangulate survey results with complaints
- Free text vs quantitative feedback – free text often easier to understand/action
- No clear procedures for acting on feedback
- Sharing of feedback variable
- Nothing ever changes as a result of surveys
- Emotional reactions

ACCEPTABILITY

PATIENTS

- Inconvenient
- "Over surveyed"
- Surveys too long
- Not completed properly
- Concerns e.g. language barriers

GPs/PRACTICE STAFF

- Surveys part of patient-centred care
- Politically driven
- May have perverse impacts if used for performance management
- Leading questions
- Forced to do it e.g. revalidation

UTILITY?

- Mixed attitudes towards surveys
- YET practice staff engage with and do not dismiss feedback
- Practice staff rarely able to coordinate effective and sustained changes as a result of feedback, particularly around more challenging areas such as clinician-patient communication
- Sustained research and practice effort required to support translation of feedback into changes BUT feedback alone unlikely to support/stimulate change

Figure 17. The 'utility index' of patient experience surveys in primary care – perspectives of practice staff

Reproduced, from Boiko, O., Campbell, J. L., Elmore, N., Davey, A. F., Roland, M., & Burt, J. (2015) [277]

Drawing on both our work and others' work, we suggest that the notion that survey feedback alone will stimulate major changes in care is unrealistic. [76, 189] Whilst we saw evidence of changes to minor modifications such as car parking, décor, and (somewhat more challengingly) appointments systems, issues such as the management of GPs with evidence of poor communication skills, or responding to other 'inter-personal' aspects of professional practice, appeared harder to tackle. Whilst patient experience will no doubt be improved by making general practices more accessible and more pleasant, significant aspects of experience linked to better clinical outcomes, including the quality of nurse- and GP-patient communication and trust and confidence in clinical staff, risk being left outside the focus of improvement work undertaken by practice staff.

There are six dimensions of the Utility Index (reliability, validity, cost, feasibility, educational impact and acceptability) which may determine the potential utility of an intervention, including patient experience survey feedback. All have relevance for how general practice staff view the current role of patient surveys:

Our identification of issues with the credibility of surveys, and difficulties in the interpretation of feedback, echoes previous work from health settings which suggests widespread scepticism about the robustness of patient surveys. [111, 184, 190, 191] Practice staff were more likely to view results positively if their scores were stable over time, were above average, and corroborated other sources of feedback such as complaints and compliments.

Whilst respondents felt national patient surveys were feasible, there were concerns about the challenges of undertaking local practice surveys. Issues included the time taken to undertake such work and how best to ensure that the conduct of in-house surveys was robust. There were also mixed attitudes about the cost-effectiveness of national survey programmes, in part due to the perceived difficulties in acting on feedback.

We identified a consistent lack of impact of surveys at practice level, driven by factors including an absence of coordinated action and difficulties in interpreting survey feedback. [192, 193] Benchmarking data was seen to be useful, although it was not always easy to make sense of. [65] Likewise, practice staff welcomed free text comments from patients as providing more specific information about their opinions. [194, 195] Most commonly, when change did happen, survey findings were only one of the prompts to address an already-acknowledged problem. Changes, however, usually focussed on service organisation or facilities and not on individual practitioner behaviour. There remains little evidence that patient feedback alone has any impact on the behaviour or skills of medical practitioners, with a number of trials having little demonstrable influence on subsequent patient feedback. [76, 189, 196] The provision of facilitated feedback of results may be more effective in engendering engagement and action, as suggested by recent evidence from secondary care. [197] However, the emotional toll of negative patient feedback on staff is also relevant here: staff reported how disheartening it could be to receive consistently poor comments. The potential to see patient feedback as threatening and harmful is an additional barrier to acting on such data, and further suggests the potential for facilitated reflection in assimilating feedback.

Practice staff worried that an endless cycle of surveys were inconvenient and burdensome for their patients. Nevertheless, surveys appeared to be broadly accepted as part of the new paradigm of patient-centred care, and were broadly welcomed in that role. However, lingering concerns over the linking of patient feedback to pay-for-performance and the external imposition of surveys on general practice tempered the acceptance of current surveying practices, particularly for GPs.

Taken together therefore, we suggest that key drivers of the gap between conducting surveys and implementing changes relate to the difficulties of practice staff in trusting and making sense of survey findings, coupled with a lack of support for identifying and making changes to practice.

Implications for practice

Whilst practice staff predominantly view feedback from patient experience surveys as a mechanism for affirming good or detecting poor service delivery, the current direction of policy targets the aspiration of providing evidence to inform changes in practice (a quality improvement mechanism). A question remains therefore as to how patient experience survey data can become a key driver of service improvement. Evidence suggests that securing feedback alone is insufficient to engender change [189] especially as practices are left responsible for developing their own implementation mechanisms. GP contractual arrangements prior to 2009 offered incentives to primary care practices to discuss the findings of patient feedback surveys with patient representatives, for example through the use of patient participation groups. Although now withdrawn, such an approach may have merits in facilitating change, as well as acting as a means of responding to the need for active patient and public participation in informing the design and configuration of services.

Recent work in secondary care highlights the potentially significant contribution of facilitators in enabling staff to review survey results and to act on them. [197] Within primary care, such initiatives are lacking. Practice staff need to be supported to reflect on patient feedback; this will need dedicated resources on top of those committed to collecting patient experience data. Quality assurance of survey development, data collection, and reporting of results is of vital importance if the findings of surveys are not to be dismissed out-of-hand on the grounds of credibility, or to become the subject of discussion aimed at diverting rather than promoting action and change.

Where surveys highlight the need for change, formal processes for planning and delivering change are required. In the current climate of scarce resources, a commitment to developing patient experience surveys as quality improvement mechanisms would therefore displace other competing priorities, and managers and practitioners must be realistic about what can be achieved. However, until then, it is our view that the full potential of patient feedback will not be achieved.

Conclusions

We have identified a number of key reasons for the gap between the receipt of patient feedback and the decision to act on that feedback. Addressing key concerns of primary care providers across all aspects of patient surveys – reliability, validity, cost, feasibility, impact and acceptability – and supporting them to reflect on the meaning of such data will be important if we are to draw on such evidence in quality improvement programmes. Alongside this, there is a need to develop a realistic understanding of where surveys may be expected to drive change, and where they may not be expected to do so.

Chapter 8. Attitudes to receiving feedback from patient experience surveys: interviews with GPs

Abstract

Background

To date, little research has focused upon doctors' attitudes to patient experience surveys which give them personalised feedback. Whilst national surveys, such as the GP Patient Survey, report results at a practice level, GPs are additionally required to reflect on individual-level patient feedback for the purposes of appraisal and revalidation. This chapter examines doctors' perceptions of patient experience surveys, and the receipt of personal feedback from these, in primary care settings.

Methods

We analysed data from 21 interviews conducted with GPs across 14 practices. Participants were sampled from doctors who had participated in our patient experience survey (reported in Chapter 9) and had recently received individual-level survey feedback.

Results

GPs expressed commitment to incorporating patient feedback in quality improvement efforts. However, they also expressed negative views about the credibility of survey findings and patients' motivations and competence in terms of providing feedback. As a result, they found it challenging to make sense of and take action as a result of the feedback they received from patient experience surveys.

Conclusions

GPs' ambivalence towards patient experience surveys is likely to limit their impact on the success of quality improvement initiatives. In response, this chapter highlights the need for initiatives to address doctors' concerns about the credibility of surveys.

Introduction and rationale

A number of recent policy initiatives have emphasised the utility of patient feedback for quality improvement. [35] In the UK, a series of initiatives have established and expanded the role of patient experience surveys in the NHS, leading to the recent NHS Outcomes Framework, [198] which features patient experience as one of five key domains upon which NHS performance is judged. In addition to national surveys, such as the GP Patient Survey, numerous surveys of various kinds are undertaken at the local level by healthcare providers. [199] In 2012, the General Medical Council introduced a revalidation programme requiring individual doctors to collect patient feedback on the care they provide. Such feedback is subsequently used as supporting information in a five-yearly procedure through which doctors 'revalidate', i.e. retain their licence to practise, and is also intended to facilitate reflective improvements in the quality of individual doctors' practise. [129] Nevertheless, most national survey programmes continue to be conducted and reported at the organisational level. Likewise, existing research has tended to focus on doctors' engagement with reports of patient experience at the level of the hospital ward, primary care practice, or similar organisational units within primary or secondary care.

Existing research highlights the importance that doctors place upon patient experience in principle, and the potential for positive improvements based on patient feedback. [111] This body of work has also explored challenges surrounding the incorporation of patient feedback into medical practice. Doctors commonly express a range of negative views about the plausibility of survey findings, including concerns about: sample size and representativeness; respondent bias and subjectivity; reliability and validity of survey instruments; lack of clarity on the purpose of surveys; contextual sensitivity; and the challenges of interpreting patient feedback when lacking contextual information, with numerical scores viewed by many doctors as 'a simplistic reduction from a complex range of factors' [184] (see also [42, 111]). These challenges relate to longstanding critiques of quantitative surveys, which highlight issues such as the lack of self-evident meaning in numerical findings (see e.g. Williams [200]) in a range of contexts including special educational services and healthcare provider performance. [201, 202] These and other concerns have tended to limit the impact of patient feedback in terms of quality improvement. [199]

Many of the challenges associated with patient experience surveys relate to standard features of survey administration, and so are also likely to be relevant to surveys administered at the individual doctor level. With some exceptions (e.g. [65]), few researchers have focused directly upon doctors’ engagements with patient experience surveys at the individual doctor level. While such engagements are largely unexplored, they are of considerable significance given the well-established role of patient experience surveys in contemporary healthcare (and the NHS in particular), and the recent introduction of mandatory individual doctor-level surveys.

This chapter draws upon qualitative data to explore attitudes towards patient survey feedback on the part of individual GPs. By exploring attitudes towards plausibility of surveys, this chapter demonstrates the generally contested, problematic, and inconsistent nature of doctors’ current engagements with patient experience surveys, and points towards the need for additional investment in training and relevant resources.

Changes to study methods from original protocol

The aim of this strand of work, as stated in the original protocol, was:

“To understand how general practices respond to low patient survey scores, testing a range of approaches that could be used to improve patients’ experience of care (aim 1).”

The interviews reported here took place alongside the focus groups with practice staff, reported in Chapter 7. In our original application, we set out plans to interview each doctor in between five and eight low-scoring practices. These interviews would cover their accounts of what contributes to their practice score, considering their recent GP Patient Survey feedback. However, as with the focus groups, the conduct of our own patient experience survey (reported in Chapter 9) at individual-doctor level meant that we were able to feedback to GPs their own patient experience scores. Interviews thus considered attitudes to both practice- and individual- level feedback. We also altered our sampling strategy, deciding instead to incorporate a wider range of practices (fourteen practices) to reflect a greater diversity of practice cultures within which the GPs were working.

Methods

Data collection

We conducted 40 semi-structured face-to-face interviews with GPs in practices across Cornwall, Devon, Bristol, north London, Bedfordshire, and Cambridgeshire. These practices were part of a larger group of 25 practices participating in our patient experience survey (see Chapter 9 for details of sampling, recruitment and survey conduct). From the sample of 25 practices, two doctors were interviewed from practices with low GPPS scores and one from each medium- and high-scoring practice. Individual GPs were identified randomly within each practice and approached one by one for consent to participate. Each GP had received an individual report from our patient experience survey, focused on patient responses to communication items and including summary statistics and free text comments. An interview topic guide was developed in light of existing literature to focus on individual-level patient experience surveys, and revised in relation to policy changes on revalidation which occurred during the conduct of the study. Interviews lasted between 20 and 60 minutes.

For the purposes of this report, we excluded 19 interviews conducted with GPs prior to the introduction of revalidation in December 2012, as this changed the nature of the topic guide and issues covered in the interviews in relation to the conduct and implications of individual doctor-level patient surveys. We thus include data from 21 GP interviews conducted across 14 practices.

Data analysis

Interviews were digitally recorded with written consent and transcribed verbatim. NVivo v.10 software (QSR International Ltd, 2012) was used to organise and categorise the data. Transcripts from four GP interviews (not included in the final analysis) were used to develop an initial coding framework, which included 44 codes grouped into headings including survey experience and survey-related change. A thematic analysis approach was used, [203] involving six distinct stages: familiarisation with the data; generating initial codes; searching for themes; reviewing themes; defining and naming themes; and producing a final analysis, which was discussed among the research team before being revised and finalised.

Results

Dimensions of ambivalence

Our analysis found that GPs demonstrated profound ambivalence regarding the purpose and plausibility of patient surveys, leading to complex, varied, and problematic engagements with patient feedback. The Oxford English Dictionary defines ambivalence as ‘having mixed feelings or contradictory ideas about something’, a definition that was interpreted in this study as a spectrum from mixed feelings about something to holding ideas that directly contradict each other. Two main dimensions of ambivalence were identified. The first relates to doctors’ views of patients’ motivations and competence as responders in surveys. The second relates to doctors’ views of surveys from the perspective of enabling quality improvement (or otherwise) – views that may diverge from what is intended by the managers responsible for introducing and administering surveys.

Interviewees rarely situated themselves consistently with regard to these two dimensions of ambivalence; indeed, it was common for GPs to express inconsistent and contradictory views on both dimensions of ambivalence, often within the same interview (see following sections). Consequently, while some themes (e.g. a greater emphasis on negative rather than positive views of patients) were more to the fore than others, ambivalence is the dominant and unifying feature of the findings in this area.

Patients and surveys

GPs emphasised the centrality of the doctor-patient relationship and the utility of receiving feedback from their patients. For example, one interviewee described the doctor-patient relationship as an ‘adult to adult’ relationship in which patients know more about some things than doctors, and in which doctors need to listen to patient feedback:

> *the only way you’re going to know whether you’re doing your job properly… it’s listening to what the patients are telling you [in their feedback]’*
>
> *GP4*

Against this backdrop, many GPs discussed patients’ motivation and competence to provide feedback in more detail. One GP discussed how patients’ feedback showed that they were reflecting in depth upon their experience before communicating it through free text comments:

> *They’re… thinking “Well, actually, what do we think of the [practice]?”… rather than just at the time when they’re desperate for an appointment and frustrated, you know, to think actually… what things at the [practice] do they actually value.*
>
> *GP9*

More widely, several doctors noted that patients were used to responding to surveys in other spheres of their lives, potentially (though not inevitably) increasing their willingness to provide feedback on their healthcare experiences. As such, many doctors saw patients as motivated to reflect upon and communicate their experiences (although this was also raised as a concern in terms of raising patient expectations; see following section).

Similarly, some doctors expressed the view that patients are competent to judge their care. Patients’ ability to evaluate doctors was sometimes endorsed because it aligned with the doctors’ pre-existing positive views of their own professional skill. However, despite this, many interviewees expounded fundamentally ambiguous views of patients considered as survey respondents, often combining in the same interview seemingly positive views of patients’ motivation and competence with more negative views. For example, one GP emphasised the utility of patient surveys in terms of patients’ capacity to identify specific problems:

> *I think the patient feedback is really important… You’ve got to actually listen to what are patients saying, [e.g.] they are telling us through this [feedback] that the system currently in place for booking appointments… is not working for them.*
>
> *GP2*

The same GP also stressed, however, the ways in which patients’ comments were often of little use for improving care quality, especially at the individual doctor level:

> *When I read the comments it was just a diatribe of accusations against the practice as a whole… [I]n terms of my individual practice it gives me no feedback at all… [The] majority of the comments on the appointment system and on lack of [relational] continuity [were] all on the issues that we are totally aware of.*
>
> *GP2*

Doctors often questioned patients' motivations firstly by viewing patients who provided negative feedback as doing so because they had specific grievances to express ('if they've got an axe to grind' [GP10]), and secondly, by suggesting that patients participate in surveys in order to gain leverage over doctors. Many interviewees also questioned patient competence, or their ability to provide accurate and relevant feedback. Overall, GPs advanced six principal characteristics of patients that singly and/or collectively undermined their ability to provide accurate feedback:

1. *Positive bias*: the tendency of patients to give strongly positive feedback regarding doctors, linked to patients' well-documented reluctance to criticise doctors in general and their own GPs in particular. In this context, one GP described their patients as 'quite reluctant to talk the doctor down, because we've got a good on-going relationship' (GP10).
2. *Negative halo effects:* patients ascribe negative characteristics to consultations because of other negative experiences during their visit to practices. As one GP described, patients may carry an 'initial bad experience' with the practice reception 'all the way through… into the consulting room as well… it affects all of your feedback' (GP6).
3. *Failure to understand surveys*: for example, one GP noted that 'because [patients] don't understand the questionnaire, they might tick whatever box they think; and that's the reason we don't get true results' (GP19).
4. S*ubjectivity*: several doctors emphasised that different individual patients could give different feedback despite having experienced similar consultations concerning similar medical problems. More widely, one GP highlighted patient subjectivity by suggesting that strongly negative patient feedback could 'reflect more on the person [patient] than it does on you [the doctor]' (GP6).

5. *Inability to evaluate clinical competence*: GPs highlighted patients’ inability to judge doctors’ clinical competence: as one noted, patients ‘don’t know about my clinical ability… [or] how much I know’ (GP8).
6. *‘Good doctors, bad feedback’*: doctors felt that good care may result in negative feedback because it differs from patients’ preferences. Common examples include doctors refusing to prescribe antibiotics or write ‘sick notes’ for patients with depression. GPs referred to situations in which patients were unhappy with treatments recommended (or withheld) by doctors, and often saw themselves as having a responsibility to protect NHS resources rather than pleasing patients: ‘pleasing a patient isn’t the same thing as being a good doctor… I see part of my role as a GP [as] gatekeeping NHS resources, including my own time’ (GP8)

Thus, while doctors’ views often combined positive and negative views of patient feedback, negative views tended to dominate, resulting in a sceptical attitude that questioned patients’ motivations and competence vis-à-vis the provision of feedback (Figure 18).

Doctors' Attitude **Category**	**Positive**	**Negative**
Patient Motivation	Willing to take time to provide feedback Used to providing feedback in other spheres	Axe-grinding Desire to influence doctors
Patient Competence	Able to recognise good quality care/improvements	Positive bias Negative halo effects of clinic/survey experiences Unable to understand survey instruments Subjective judgements Lack of clinical knowledge Good doctor/bad feedback

Figure 18. GPs' attitudes to patients' motivation and competence

Patient experience surveys and quality improvement

This section focuses on a second dimension of ambivalence, relating to GPs' perceptions of the potential for patient experience surveys to drive quality improvement. Doctors identified benefits in reflecting upon patient feedback and encouraging competition between doctors through comparison of patient feedback scores. However, they also presented a number of concerns that undermined the potential of surveys to facilitate quality improvement. As with doctors' attitudes towards patients' feedback, the overall impression was more negative than positive.

Positive Attitudes

Doctors emphasised the potential for patient experience surveys to facilitate quality improvement in a variety of ways. One GP, for instance, emphasised that they 'actually took on board things which people were saying', since 'there's no point doing a survey… unless you're actually going to take notice of what the results say' (GP1). Numerous participants described negative feedback as having more utility for change than positive feedback. Furthermore, a number of doctors discussed the potential for quality improvement to be driven by doctors' competitiveness with regard to colleagues' performance and/or benchmarked data (i.e. data supplied alongside comparative figures for comparable surveys undertaken in the past or elsewhere). One GP, for example, noted that surveys are:

> *all about comparing yourself with other GPs who do the same job… Because, I think, you want to know that you're in the best group, compared with other people*
>
> *GP5*

Overall, interviewees saw the potential for survey-based quality improvement in three main areas:

(a) *Reminders of core proficiencies*, especially communication skills and basic tasks such as introducing themselves to patients and ensuring that patients are satisfied with the consultation before they leave. Several doctors remarked on the utility of repeated surveys for highlighting the importance of such issues; one GP said, 'I think it flags up … the initial consultation tips that you *think* you do that perhaps you don't always' (GP5).

(b) *Reinforcements of known problems (and providing evidence to support change),* often at the practice level: ‘the [survey] was useful because [it] really reinforced the impressions that we were beginning to form as… colleagues, and it was a bit more evidence that we could actually say, “Well, look, this isn’t personal, because look at this, and this is random and anonymised data coming in”’ (GP15)

(c) *Unexpected issues documented in free text comments*. These were often seen as providing more useful material for reflection and change than numerical feedback, which was seen as overly positive about the care patients had received. Thus one GP stated that ‘I actually took more from the free text comments… because I think the figures were… all pretty good really… [R]eading through the comments I think is really quite helpful… just having it there makes you think about it and think “Well, why do I do that?”’ (GP1).

Negative Attitudes

Doctors’ positive attitudes towards the potential of patient experience surveys in facilitating quality improvement, noted above, were paralleled and undermined by a plethora of sceptical views. For interviewees, this led to an ambiguous but overall decidedly negative picture in which the value of surveys for quality improvement purposes was placed in severe doubt, in line with preceding research in other fields that emphasises the challenges involved in interpreting survey data. [201, 202] As well as negative views of patient motivations and competence, outlined above, GPs added several more reasons for discounting surveys as quality improvement tools. Broadly, these concerns fell into five categories:

(a) *Concerns about the validity and reliability of surveys* on the basis of factors including low response numbers, biased samples, and problematic administration methods. GPs expressed concern about response numbers despite having high numbers of respondents for their individual feedback (with a mean of 71, double the usual number required for adequate reliability). One GP linked what they saw as low response rates to patients’ fatigue regarding surveys: ‘There is a little bit of questionnaire overload… And I think it’s reflected in a very poor response rate’ (GP2).

(b) *Difficulties surrounding interpretation*, including the separation of statistics from free text comments and thus the difficulty of interpreting patients’ rationale for specific responses in a given survey. As one GP remarked, ‘if there was a problem there [in the numbers] I’d look towards addressing that, but I couldn’t really find a comment which was associated with that … so I found it quite difficult’ (GP1). As research has found in other contexts, [197] feedback presented to healthcare professionals without expert facilitation can be difficult to interpret and act upon.

(c) *Issues of context*. Doctors raised concerns about specific features of clinical encounters or patient characteristics which could influence patient feedback and thus undermine the value of patient feedback as a foundation for quality improvement. For example, some GPs who worked in deprived areas felt that surveys did not take sufficient account of the possibility of some population groups giving systematically more negative feedback than other groups: ‘sometimes I think you have a survey and I don’t think it’s a true reflection of where you are, your demographics. And I think that can be a problem’ (GP11).

(d) Anxiety about negative feedback. A number of GPs discussed actual or potential anxiety arising as a result of negative feedback. This could problematize doctors’ engagements with survey findings, impact on their confidence, and make them less likely to adopt a positive and constructive attitude towards improving their care. One GP, for instance, described feeling upset and worried following negative feedback – feelings that were shared by many other GP interviewees: ‘I find it quite difficult, because I’ll always take it quite personally’ (GP3).

(e) The risk of raising patient expectations. The fifth and final area of concern relates to the risk of raising patient expectations through surveys by introducing a consumerist element previously more associated more with customer relations than medicine. As one GP noted, ‘it’s like TripAdvisor, everything, everybody’s being rated’ (GP8). As several doctors noted, it is not always possible to meet these rising expectations, especially with regard to resource-related issues such as out-of-hours appointments; consequently, surveys may encourage patients to expect changes that are impossible to implement in practice, leading in turn to negative patient feedback. Thus, if quality improvement is evaluated at least in part on the basis of patient experience surveys, then surveys themselves may render evidence of improvement less likely.

Overall, negative views of the potential contribution of patient surveys to quality improvement agendas dominated the findings (Figure 19).

Positive	**Negative**
• Value of reflecting upon patient feedback • Value of competition between doctors on the basis of survey feedback • Reminders of core proficiencies • Reinforcements of known problems (and providing evidence to support change) • Unexpected issues documented in free text comments	• Discounting of patient motivations and competence • Concerns about the validity and reliability of surveys • Difficulties surrounding interpretation • Issues of context • Anxiety about negative feedback. • Risk of raising patient expectations

Figure 19. Doctors' attitudes to patient experience surveys as quality improvement tools

Discussion

This study explored GPs' engagements with patient experience surveys, and our findings show that that they express mixed and contradictory opinions, despite long-standing experience of such surveys. We have discussed doctors' views with regard to two key dimensions of ambivalence: patients as responders to surveys, and the potential of patient feedback to facilitate quality improvement agendas. Interviewees' vacillation between different standpoints problematized attempts to generate a clear impression of engagement with patient experience surveys. Nevertheless, it is possible to draw some important conclusions. Whilst GPs endorsed patients' motivations for participating in surveys and their competence to provide accurate and relevant feedback, these notions were outweighed by the numerous ways in which doctors emphasised what they saw as patients' questionable motivations and lack of competence vis-à-vis surveys. Consequently, doctors appear to view patients, as survey respondents, in a deeply ambiguous fashion – i.e. as being simultaneously competent and incompetent at evaluating doctors, as being both accurate reporters of experience and inevitably biased commentators, as disinterested contributors to quality improvement and axe-grinders. Likewise, while participants appeared to emphasise the potential utility of patient feedback for quality improvement, they also presented numerous factors which individually and collectively undermined this agenda. Overall, GPs' engagements with patient experience surveys were highly contested, problematic, and inconsistent.

From a GP's perspective, surveys themselves exhibit varied properties and capacities in multiple dimensions, including the different ways patients are perceived to interact with survey instruments, the different purposes for which surveys can be undertaken, and the different conceptualisations doctors can generate about them, reflecting the wider challenges inherent in interpreting numerical data. [197] Additionally, the varied ways in which survey feedback is disseminated in different settings generates another tier of properties: a feedback report that is emailed to a doctor with no benchmarking or interpretative guidance is effectively a different kind of feedback from a benchmarked report discussed with a facilitator. As a result of these varied properties, ambiguity is a strong feature of surveys as currently administered. GPs appeared to make sense of this ambiguity by drawing on their identities and frames to arrive at a widely shared yet 'internal' ambivalence. In this context, we understand 'internal' ambivalence as a kind of ambivalence that takes place not so much *across* different doctors (though this was seen at times) but rather *within* doctors, such that individuals tended to express multiple and mutually contradictory ideas. From this perspective, doctors appeared to consider more than one interpretation of patient experience surveys as plausible at the same time.

Nevertheless, doctors did not see all interpretations as equally plausible. As discussed above, they tended to settle on negative views of patients and of patient experience surveys, thus undermining the potential for reflective change and quality improvement in response to patient feedback (in line with previous research [111, 184]). The numerous specific reasons that doctors gave in support of their standpoints - ranging from patients' lack of clinical expertise to surveys' lack of contextual sensitivity (see Table 24, below) suggest that plausibility in this context is a complex, multi-layered, and largely 'negative' phenomenon. As such, patient experience surveys can be seen as an important instance of a wider problematic identified by May et al [204]: 'what to do with the patient's subjective experience of illness, and how to connect it with medical knowledge and practice.' Research in other domains, such as teachers' responses to pupils' feedback, illustrates that this problem is not specific to healthcare contexts, but, rather, characterises more universal responses to feedback and criticism. [205]

Implications for practice

A particular aim of this strand of work was to consider what approaches might be used to improve patients' experience of care. Our findings suggest that some basic steps are first required to improve the credibility of survey findings in the minds of GPs and increase their engagement with them. While the 'internal' ambiguity exhibited by GPs – i.e. the co-existence of positive and negative views of patient experience surveys – demonstrates the problematic nature of doctors' engagements with patient experience surveys, it also suggests the possibility of positive change in the future by building on some of the positive views that doctors already hold regarding patients and surveys. In the patient survey context, opportunities exist for managers and lead clinicians to engage in processes aimed at strengthening the plausibility of patient feedback surveys. For GPs to see quality improvement on the basis of patient feedback as plausible, these findings suggest they would need to be persuaded simultaneously of: patients' evaluative competence and disinterestedness; the possibility of interpreting feedback meaningfully; the ability of survey instruments to take account of contextual factors; the provision of support for doctors receiving negative feedback; and assurance of measures to limit the risk of raising patient expectations (except where it is intended to raise patient expectations). In each of these arenas, as presented in Figure 19, potential exists for measures to be taken. By doing so, relevant stakeholders can help to shape GPs' engagements with patient surveys in more positive directions.

Table 24. Plausibility of patient experience surveys: limiting factors and potential solutions

Factors inhibiting plausibility of interpretations favouring quality improvement		**Solutions to increase plausibility**
Views of patients	Not dispassionate evaluators	Facilitate doctors' personal engagement with patients; training for doctors regarding psychometric bases of validity
	Incompetent evaluators	Facilitate doctors' personal engagement with patients; include clearer instructions to patients on survey instruments
Views of surveys	Difficulties of interpreting feedback	Provide facilitated feedback for individual doctors/groups of doctors, embedded within wider local change programmes; provide additional information on feedback material (e.g. benchmarking data)
	Lack of contextual sensitivity	Explore potential for development/ validation of tailored survey instruments for different care settings
	Anxiety regarding negative feedback	Provide support for individual doctors concerned about negative feedback
	Risk of raising patient expectations	Limit frequency of survey administration to minimum necessary, except where raising patient expectations is intended

Conclusions

This chapter has explored the ambiguities in GPs' attitudes to patient experience surveys, and has focused on the plausibility of survey findings. While policy developments over the past decade have increasingly emphasised the importance of patient experience surveys in terms of quality improvement, these findings suggest that this agenda faces significant challenges in terms of doctors' inconsistent and highly critical engagements with patient feedback. GPs discount patients' motivations and competence at the same time as emphasising patient-centred care, and undermine the potential for survey-based quality improvement while also highlighting the importance of patient feedback. GPs demonstrated complex and ambivalent attitudes towards the plausibility of patient experience – attitudes that are likely to constrain the potential impact of patient experience surveys on care delivery. In response, we highlight the need for initiatives on the part of managers and lead clinicians in order to address doctors' plausibility concerns.

Chapter 9. Understanding high and low patient experience scores: analysis of patients’ survey data for general practices and individual GPs

This material is reproduced, from Roberts, M. J., Campbell, J. L., Abel, G. A., Davey, A. F., Elmore, N. L., Maramba, I, et al. Understanding high and low patient experience scores in primary care: analysis of patients’ survey data for general practices and individual doctors. 2014; BMJ. 349:g6034 [118] *under the terms of a Creative Commons licence. Articles published under the CC-BY permit unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.*

Abstract

Background

There is increasing interest in collecting and, potentially, publishing performance data at an individual practitioner level, in part to enable patients to make informed choices about their care provider. However, UK general practice performance data remains at the practice level, potentially masking important differences between individual practitioners. The aim of this strand of work was to determine the extent to which practice-level scores mask variation in individual performance between doctors within a practice. Additionally, we aimed to determine the test-retest reliability of core items derived from the GPPS.

Methods

Patient experience surveys were sent to patients who had recently had a face-to-face GP consultation in one of a stratified sample of GP practices. In addition, a subsample of patients returning questionnaires were sent a retest questionnaire.

Results

Main survey: 7721 patients consulting one of 105 GPs across 25 practices returned questionnaires (response rate 50.9%). The proportion of variance in communication scores attributable to differences between doctors (6.4%) was considerably more than that attributable to practices (1.8%). Higher-performing practices usually comprise only higher-performing doctors, but lower-performing practices may include doctors with a wide range of communication scores. Test-retest: 348 patients consulting one of 20 GPs from 5 practices returned a retest questionnaire (response rate 58.3%). Percentage agreements for categorical items between test and retest ranged from 66% to 100% (kappa coefficients range 0.00-1.00). The intra-class correlations for ordinal items averaged 0.67 (range 0.44-0.77).

Conclusion

Aggregating doctors’ communication scores at practice level can mask considerable variation in individual doctor performance, particularly in lower-performing practices. Most of the items derived from the GPPS have moderate to almost perfect reliability, with performance-related items achieving substantial reliability.

Introduction and rationale

Public reporting of performance at either provider or individual level is increasingly becoming the norm in healthcare. The approach is proposed to increase accountability, transparency and public engagement, [78, 79] It is proposed that offering users the potential to compare their primary care provider with other similar providers may enable patients to make more informed decisions about their care, although evidence in this area is limited. [206] The major source of compiled and published patient feedback and GP practice performance scores, the English General Practice Patient Survey (GPPS), is currently available on websites such as NHS Choices (www.nhs.uk) and Compare (http://www.locallyhealthy.co.uk/perf). [187] Making such data publicly available may provide a comprehensive overview of NHS primary care performance, although it is not without controversy. [5]

The GP Patient Survey collects patients' views on the quality of care they receive from their local GPs, dentists and out-of-hours doctor services. It includes a series of items on the inter-personal skills of the last GP they saw at their practice (within the previous six months). A significant limitation of GP Patient Survey ratings, however, is that items relating to the doctor-patient relationship are reported at practice level, possibly masking considerable performance variation amongst individual GPs within that practice. Aggregation of ratings may offer both inaccurate reporting of patient views of individual doctor performance, and little scope for reflection on the part of GPs about their personal strengths and weaknesses. Current indicators may consequently fail to provide users, providers or commissioners with an accurate assessment of performance within a practice.

Changes to study methods from original protocol

The aim of this strand of work, as stated in the original protocol, was:

"To estimate the extent to which aggregation of scores to practice level in the national study masks differences between individual doctors (aim 2)."

In this chapter, we address two of the original four main objectives for the study:

- To provide scores for individual doctors, to allow us to estimate the extent to which aggregation of scores to practice level in the national survey masks differences between individual doctors within practices.
- To explore the extent to which patient responses to items used in the GP Patient Survey show stability over time (7-10 days)

In our original application, we envisaged using this strand of work to additionally (a) identify patients for cognitive interviewing (as reported in Chapter 2) and (b) identify patients of South Asian for our work on variations in patient experience in minority ethnic groups (as reported in Chapters 2, 5 and 6). In practice, these two objectives were moved to stand-alone studies as a result, in part, of our switch of survey distribution from face-to-face, as originally planned, to postal mode. Pilot work showed that the distribution of a post-consultation survey by our research team in participating practices over a defined data collection period risked a high proportion of “missed” eligible patients and a high burden to research staff to accomplish this for over 7,000 respondents. Instead, after further pilot work, we undertook a postal survey, working with practices to identify patients with face-to-face consultations within the previous three weeks: see below for full details.

Methods

Twenty-five general practices from Cornwall, Devon, Bristol, Bedfordshire, Cambridgeshire and North London were invited to participate. The aim was to recruit 15 practices scoring in the lowest 25% of all practices in the 2009/10 GP Patient Survey on a composite case-mix adjusted score for the doctor-patient communication items in the questionnaire, and five practices scoring in each of the middle and highest quartiles (i.e. 37.5th-62.5th or above the 75th percentile). Linear regression models were used to adjust for patients’ age, sex, ethnicity, deprivation score, and self-rated health for case mix. Practices had to have at least two registered General Practitioners (GPs), and the sample was stratified by practice-level communication score and by GP head count, deprivation index and geographical location. All GPs (working at least four sessions per week and not trainees, short term locums or currently on extended leave) within each practice were required to be willing to participate. Practices were approached in a randomised order until the quota for each stratum was achieved.

Data collection took place during October 2011 to June 2013. A list of face-to-face GP-patient consultations conducted three weeks prior to the specified date was extracted from electronic practice records. Practice staff screened lists for recent deaths, terminal illness, under 18s and mental incapacity. Once the extracted list was screened the remaining patients were sent a patient experience survey accompanied by a letter from the practice, a study information sheet and a prepaid envelope. Repeat consulting patients were sent one questionnaire only, which related to their most recent consultation at time of extraction. Non-responders were sent one reminder within three weeks of the initial mail-out and questionnaires returned up to 100 days after the initial mail-out were accepted.

Fifty completed questionnaires were judged sufficient for obtaining reliable mean communication scores for comparable patient feedback instruments. [207, 208] The survey cycle was thus repeated until 50 completed questionnaires for each participating GP were received or until three cycles were complete. Patient consent to take part in the study was inferred by receipt of a questionnaire.

Questionnaire used in this study

The questionnaire was based on the instrument used in the national GP Patient Survey, and asked patients about access, waiting times, opening hours, continuity, interpersonal aspects of care, and also demographic details, including self-rated health. The Improve advisory group had particular input into the design of the study materials, including the questionnaire. Patients were asked to recall and report on a consultation with a specified GP on a specified date (corresponding to details extracted from practice records) when completing seven communication items and one confidence and trust item. A mean communication score for the GP from each respondent was calculated from the seven communication items (questions 22a-22g, see Appendix 22 for the full questionnaire), for patients providing four or more informative responses.

Test-retest reliability

GPs within the five participating practices with the highest response rates from the initial mail-out were selected to take part in the retest phase. Patients returning the test phase questionnaire within three weeks of mail-out were sent a retest pack, containing a differently-coloured questionnaire, covering letter and an information sheet. Only retest questionnaires returned within 4 weeks of their initial mail-out were accepted. The gap between completion of the first (test) questionnaire and of the retest questionnaire varied between 3 and 49 days; the gap between the consultation and completion of the retest varied between 30 and 76 days.

Analysis

Main analysis

The gender balance, proportion of doctors who trained in the UK and mean time since registration in the practice sample, together with questionnaire response rates and intervals between patient consultations, mail-out and receipt of questionnaires were described. A two sample *t*-test was used to test whether intervals between consultation and mail-out were associated with questionnaire responses.

In our study design, groups of individual patients' scores are associated with (nested within) individual GPs, and groups of GPs are associated with individual practices. Although some variance in patients' scores could be attributed to individual experiences, some of the variance was likely to be attributed to GPs, as well as to other aspects of the practices (e.g. reception staff, opening hours). Three-level mixed-effects hierarchical linear models were used to estimate the extent to which variance for each outcome measure was attributable to the differences between practices, between doctors within each practice, and to the patients and other residual scores.

The models were adjusted for four self-reported patients' attributes shown to be important predictors of reported patient experience: gender, age, ethnicity, and self-reported health status. [209] The practice, doctor, and patient-related variance components from each model were expressed as percentages of the total variance. The "best linear unbiased predictors" of the practice and doctor effects were used to provide estimates of the mean score for each doctor on each of the outcome measures. [210] Corresponding estimates of the mean scores for each practice were elicited from additional models, omitting random effects for doctors. The variation in GP and practice mean scores were described and simple correlation analysis investigated the association between the practices' mean score and the within-practice standard deviation of the GPs' mean scores. The variance components from each model were used to estimate the number of patients' scores per doctor needed to achieve a reliability of at least 0.7 or 0.8 for the doctor's mean score (see Appendix 23 for formula). Whereas a reliability of 0.8 or higher is desirable for moderate to high stakes assessments, [211] a threshold of 0.7 was regarded as acceptable in patients' assessments of doctors' performance in some contexts. [212] Stata v.10.1 (StataCorp, Texas, USA) was used for data analysis.

Analysis of test-retest reliability

The response rate and response timings for both test and retest phases were described, and the demographic profiles of three groups of patients were compared: those who were sent but did not return a test questionnaire within three weeks of mail-out (not eligible for retest), those who were sent but did not return a retest questionnaire within four weeks of mail-out, and those who returned both test and retest questionnaires within the deadlines. The proportions of non-response by patients eligible to answer each of the 54 separate items were compared between the test and retest phases using chi-squared tests with a Holm-Bonferroni correction for multiple comparisons. [213] For the 33 categorical response items, the test-retest reliability was measured using raw agreement rates and Cohen's Kappa statistic. [214] Integer scores were assigned to meaningful response options (ignoring 'Don't know' or 'Not applicable' options) for the 21 ordinal response items, and ICCs were calculated. Both ICCs and the Kappa statistics were interpreted as follows: <0.00 was judged 'poor', 0.00-0.20 'slight', 0.21-0.40 'fair', 0.41-0.60 'moderate', 0.61-0.8 'substantial' and 0.81-1.00 'almost perfect'. The mean score on each item in the test and retest phases were calculated, and paired sample t-tests using the Holm-Bonferroni correction were used to test possible changes in the mean scores. Data analysis was conducted using SPSS version 18 (SPSS 2009).

Results

Of 59 practices initially approached, six were ineligible, nine declined participation, and 19 did not respond by the time the quota (n=25) was achieved. 105 doctors participated (mean 4.2 (range 2-8) per practice), 46% were female, 80% trained in the UK, they had an average of 19.5 years (range 4-38) experience since registration with the General Medical Council (Table 25). Table 26 provides an overview of responders' demographics. The mean interval between the patient's consultation date and questionnaire mail-out was 16.6 days (SD 6.0) and there was no evidence that the interval length was related to the likelihood of a completed questionnaire being returned (two sample t test, P=0.157). The overall questionnaire response rate was 50.9% (7721/15 172), range 23.6%-80.7% for individual GPs, and 24.1%-75.5% for practices. 92 out of 105 (87.6%) GPs achieved 50 returned questionnaires. The mean interval between the patient's consultation and receipt of their completed questionnaire was 35.3 (SD 15.5) days. Questionnaires with fewer than four informative responses to the seven communication items were excluded and scores for the 7429 (96.2%) responding patients were calculated, with a mean communication score of 87.5 (SD 17.8) on a 0-100 scale.

Table 25. Practice profiles and questionnaire response rates

Setting	Banding on 2009/10 GPPS communication score[1]	GP head count	Participating doctors	List size (000's)	Deprivation index[2]	Overall response rate (%)
Inner city	Low	2	2	6.9	26.6	37.9
Inner city	Low	3	3	5.1	48.5	36.8
Inner city	Low	4	4	5.1	36.6	37.8
Inner city	Low	5	4	7.8	26.1	50.5
Inner city	Low	8	6	8.7	32.4	43.5
Inner city	Middle	2	2	2.5	30.1	47.0
Inner city	Middle	3	3	5.4	13.7	67.7
Inner city	Middle	6	6	8.0	39.4	32.0
Urban	Low	2	2	3.5	15.2	71.0
Urban	Low	2	2	2.9	22.2	58.9
Urban	Low	2	2	3.2	29.6	24.1
Urban	Low	3	3	6.6	15.1	55.8
Urban	Low	4	4	4.1	18.3	59.3
Urban	Low	5	5	12.0	27.6	58.9
Urban	Low	5	5	6.0	19.3	52.6
Urban	Low	7	6	9.7	20.0	53.8
Urban	Low	8	7	16.5	14.4	45.1
Urban	Low	9	8	11.8	16.4	48.1
Urban	Middle	3	3	5.3	20.8	67.8
Urban	High	6	5	8.5	22.1	47.2
Urban	High	8	8	14.2	18.9	64.4
Rural	Middle	5	4	5.1	23.1	60.5
Rural	High	3	2	2.4	18.9	49.8
Rural	High	4	4	5.4	11.5	75.5
Rural	High	5	5	9.1	4.8	71.7
All		114	105			50.9

1. Low = below the 25th percentile, Middle = between the 37.5th and 62.5th percentiles, High = above the 75th percentile.
2. These scores are an average taken across the practice population and underlie the figures reported by Public Health England at http://fingertips.phe.org.uk/profile/general-practice .

Reproduced, from Roberts, M. J., Campbell, J. L., Abel, G. A., Davey, A. F., Elmore, N. L., Maramba, I, et al. (2014) [118]

Table 26. Demographic profile of patient respondents

	N (% non-missing)
Gender	
Female	4785 (62.4)
Male	2882 (37.6)
(Missing)	*54*
Age	
under 18	5 (0.1)
18 to 24	249 (3.2)
25 to 34	786 (10.3)
35 to 44	983 (12.8)
45 to 54	1150 (15)
55 to 64	1474 (19.2)
65 to 74	1550 (20.2)
75 to 84	1171 (15.3)
85 or over	299 (3.9)
(Missing)	*54*
Ethnicity	
White British	6138 (81.5)
White Irish	132 (1.8)
Any other white background	459 (6.1)
Mixed White and Black Caribbean	23 (0.3)
Mixed White and Black African	10 (0.1)
Mixed White and Asian	18 (0.2)
Any other mixed background	19 (0.3)
Asian or Asian British - Indian	169 (2.2)
Asian or Asian British - Pakistani	55 (0.7)
Asian or Asian British - Bangladeshi	71 (0.9)
Any other Asian background	72 (1)
Black or Black British - Caribbean	95 (1.3)
Black or Black British - African	161 (2.1)
Any other Black background	9 (0.1)
Chinese	45 (0.6)
Any other ethnic group	57 (0.8)
(Missing)	*188*
Health	
Poor	714 (9.5)
Fair	1827 (24.3)
Good	2502 (33.2)
Very good	1961 (26.1)
Excellent	523 (6.9)
(Missing)	194
All	7721

Reproduced, from Roberts, M. J., Campbell, J. L., Abel, G. A., Davey, A. F., Elmore, N. L., Maramba, I, et al. (2014) [118]

Twenty doctors from five practices took part in the test-retest sub-study. In the test phase 2,877 patients who had recently consulted one of the participating GPs from the five practices were sent questionnaires. Retest questionnaires were sent to 597 patients who had returned a completed test questionnaire within three weeks of mail-out and 58% (348/597) returned a completed retest questionnaire within four weeks. The mean time from mail-out to receipt of a completed questionnaire was 8.7 days in the test phase, and 10.1 days in the retest phase. There were no gender differences between test and retest respondents, but retest responders tended to be older and White British (Table 27). No significant differences in item non-response rates between the test and retest phase were found for any of the 54 items.

Table 27. Demographic characteristics of test-retest patient sample by level of study participation with P value for tests of variation across the three groups

	Patients sent but not returning a test questionnaire within three weeks of mail out.*	**Patients sent but not returning a retest questionnaire within four weeks of mail out.**	**Patients returning both a test and a retest questionnaire within the deadlines.**	**P value**
Number	2,009	249	348	n/a
Number (%) male	807 (40.1)	89 (35.7)	138 (39.7)	0.404
Number (%) White British*	404 (88.0)	204 (89.1)	326 (95.6)	0.001
Mean (SD) age in years	46.2 (18.5)	59.4 (18.8)	65.3 (15.1)	<0.001

*Ethnicity data was only available for those who returned a completed test questionnaire and responded to the ethnicity item. For patients sent but not returning a test questionnaire within three weeks of mail out, n=473; for patients sent but not returning a retest questionnaire within four weeks of mail out, n=229; for patients returning both a test and a retest questionnaire within the deadlines, n=341.

Main results

For the six outcome measures of interest, most of the variance in patient level scores was due to differences in ratings of the same doctor by different patients (Table 28). For both GP communication and trust and confidence in the doctor, the variance due to differences between doctors was greater than that attributable to differences between practices: however, the reverse was true for the other four, non-doctor-specific measures. Table 29 shows the number of patient ratings required to achieve the 0.7 and 0.8 reliability thresholds for each outcome measure, judged by authorities to represent minimum acceptable thresholds in postgraduate assessment settings. A substantial majority of doctors received sufficient scores to achieve reliable estimates of performance in communication – 103 out of the 105 GPs received at least 27 patients' communication scores, and 95 GPs received 46 or more (overall mean 71 scores per doctor).

Table 28. Percentages of variance in adjusted mean outcome scores that are attributable to practices, doctors and patients

	Source of variance		
Outcome measure	**Practice**	**Doctor**	**Patients and residual error**
Communication score	1.8	6.4	91.9
Confidence & trust	0.8	5.2	94.0
Overall satisfaction with surgery	6.0	1.1	92.9
Helpfulness of receptionists	7.3	0.5	92.2
Cleanliness of health centre	10.6	0.3	89.1
Ease of getting into building	1.9	0.4	97.6

Reproduced, from Roberts, M. J., Campbell, J. L., Abel, G. A., Davey, A. F., Elmore, N. L., Maramba, I, et al. (2014) [118]

Table 29. Number of patient ratings needed to achieve reliability of 0.7 or 0.8 for a doctor's raw and adjusted mean scores.

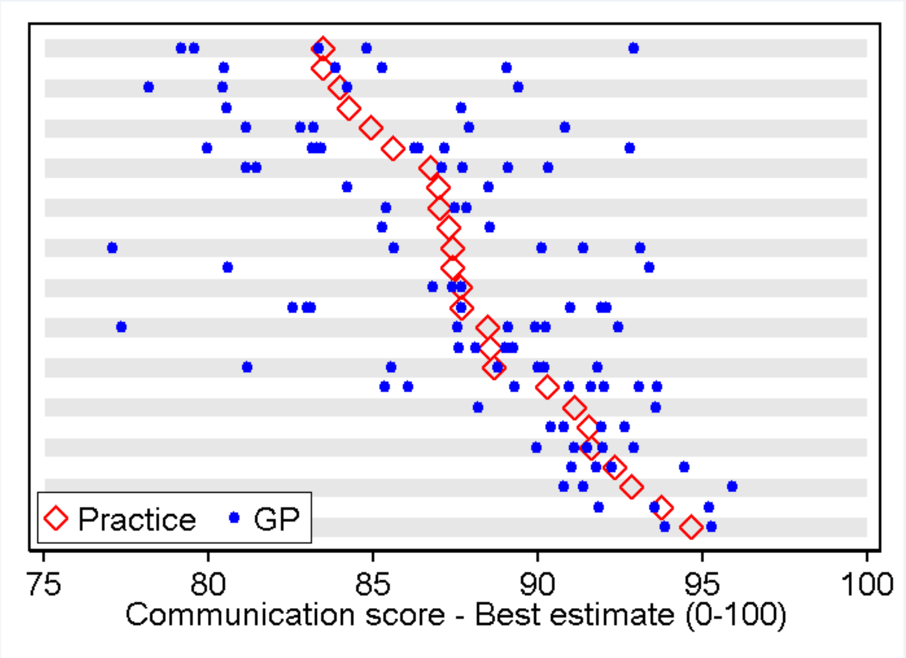
	Communication score	**Confidence & trust**	**Overall satisfaction with surgery**	**Helpfulness of receptionists**	**Cleanliness of health centre**	**Ease of getting into building**
Reliability of raw mean score						
0.7	21	30	23	25	15	78
0.8	36	51	38	42	26	133

Reliability of adjusted mean score *						
0.7	27	37	31	28	20	97
0.8	46	63	53	48	33	167

*Adjusted for patient's gender, age, ethnicity, and self-reported health status

Reproduced, from Roberts, M. J., Campbell, J. L., Abel, G. A., Davey, A. F., Elmore, N. L., Maramba, I, et al. (2014) [118]

The estimated mean communication scores for individual doctors and for practices as a whole are shown in Figure 20. This shows the extent to which the variation in mean communication scores between individual doctors (within practices) was greater than the variation between practices, and suggests that within-practice variability in doctors' scores was greater in the lower scoring practices. We conducted further analysis to confirm this: the within-practice standard deviation of GPs' mean communication scores was negatively correlated with the practice's mean communication score (Pearson's r=−0.505; P=0.010).

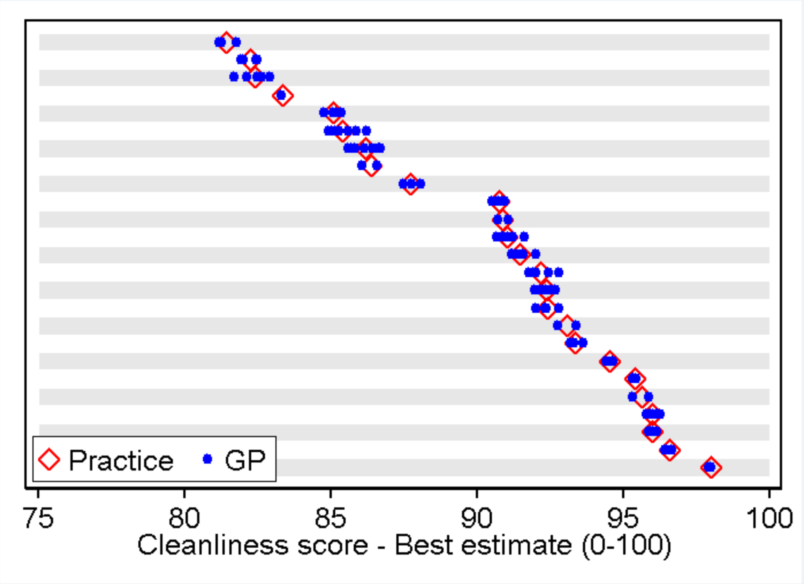


Note: Practices (n=25) are sorted by their mean communication score. Horizontal shading serves only as visual separation of results for different practices. Reliability calculations using variance components showed that achieving acceptable reliability (>0.7) for general practitioners' adjusted mean communication scores with 27 patients' scores and good reliability (>0.8) with 46 patients' scores per doctor is feasible (see appendix for formula). All but 10 of the 105 participating doctors had more than 46 scores; two received less than 27 scores (mean 71 scores per doctor). Data for these doctors was retained in the subsequent modelling, as use of best linear unbiased predictors to estimate doctors' mean scores has a "conservative" effect. Where sample sizes are smaller, estimated mean scores are drawn closer to practice mean

Reproduced, from Roberts, M. J., Campbell, J. L., Abel, G. A., Davey, A. F., Elmore, N. L., Maramba, I, et al. (2014) [118]

Figure 20. Mean communication score (best estimate) by practice and doctor

In contrast to Figure 20, Figure 21 highlights the adjusted doctor-level and practice-level mean scores for "cleanliness of the practice buildings", and demonstrates the minimal within-practice variability between GPs for this non-doctor-specific measure.



Note: Practices (n=25) are sorted by their mean score for cleanliness. Horizontal shading serves only as visual separation of results for different practices

Reproduced, from Roberts, M. J., Campbell, J. L., Abel, G. A., Davey, A. F., Elmore, N. L., Maramba, I, et al. (2014) [118]

Figure 21. Mean score for cleanliness of practice building (best estimate) by practice and doctor

Results of test-retest reliability analysis

The percentage agreement in responses to the 33 categorical items ranged from 66% to 100% (mean 88%), while the kappa coefficients ranged from 0.00 to 1.00 (mean 0.53). Only one item, relating to booking an appointment by fax, achieved a perfect agreement (kappa 1.00) (Table 30). The raw agreement rates were 80% or above for 27 of these items. ICCs for the 21 ordinal items averaged 0.67 and ranged from 0.44 for question 9 ("How easy do you find it to get into the building at this GP surgery or health centre?") to 0.77 for question 25 ("Would you recommend this GP surgery or health centre to someone who has just moved to your local area?"). The ICCs for 20 of these items (excepting 9) were above 0.6, representing substantial test-retest reliability. Mean scores in the retest phase were higher for 8 and lower for 12 of the 21 items (Table 31). After applying the Holm-Bonferroni procedure question 9 was the only item for which a significant difference was found between the mean scores in the test and retest phase (p = 0.001).

Table 30. Sample size, raw agreement (%) and Cohen's kappa statistic for the 33 categorical items

Topic / Item	N	Raw agreement	Kappa
Making an appointment			
Q1a Normally book an appointment in person	348	82%	0.63
Q1b Normally book an appointment by phone	348	95%	0.69
Q1c Normally book an appointment by fax	348	100%	1.00
Q1d Normally book an appointment online	348	99%	0.93
Q1e Normally book an appointment by digital TV	348	100%	a
Q1f Booking doesn't apply	348	99%	0.00
Q2a Prefer to book in person	348	81%	0.62
Q2b Prefer to book by phone	348	85%	0.44
Q2c Prefer to book by fax	348	99%	0.50
Q2d Prefer to book online	348	93%	0.79
Q2e Prefer to book by digital TV	348	100%	a
Q2f No preference in booking an appointment	348	98%	0.39

Access to a doctor			
Q4 In the past 6 months, have you tried to see the doctor quickly	334	82%	0.49
Q5 Were you able to see the doctor quickly	234	83%	0.46
Q6a If you couldn’t be seen quickly was this because there were no appointments	348	83%	0.39
Q6b If you couldn’t be seen quickly was this because there the times did not suit you	348	97%	0.46
Q6c If you couldn’t be seen quickly was this because the appointment was with a doctor you didn’t want to see	348	94%	0.44
Q6d If you couldn’t be seen quickly was this because the appointment offered was with a nurse and you wanted to see a doctor	348	99%	0.46
Q6e If you couldn’t be seen quickly was this because you were offered an appointment at a different branch	348	98%	0.44
Q6f If you couldn’t be seen quickly was this because there was a different reason	347	98%	0.43
Q6g Can’t remember why you were unable to be seen quickly	348	97%	0.43
Q7 In the past 6 months, have you tried to book ahead for an appointment with a doctor	339	79%	0.44
Q8 Were you able to get an appointment with a doctor more than 2 weekdays ahead	239	73%	0.40
Arriving at the appointment			
Q11 In the reception area, can other patients overhear what you say to the receptionist	339	80%	0.59
Continuity of care			
Q15 Is there a particular doctor you prefer to see	338	91%	0.68
Q17 Was your consultation with your preferred doctor	254	89%	0.55

Opening hours			
Q19a As far as you know is the surgery open before 0800	330	75%	0.46
Q19b As far as you know is the surgery open at lunchtime	309	71%	0.49
Q19c As far as you know is the surgery open after 1830	307	66%	0.47
Q19d As far as you know is the surgery open on Saturdays	309	80%	0.42
Q19e As far as you know is the surgery open on Sundays	308	85%	0.38
Q20 Would you like the surgery to be open at additional times	313	83%	0.57
Q21 Which additional time would you most like your surgery to be open	111	77%	0.49

a. Left unticked by 100% of respondents in both phases. Kappa cannot be calculated.

Table 31. Sample size, ICC (95% confidence interval), mean test-retest difference (95% confidence interval) and associated P value for the 21 ordinal response items

Topic / Item	**N**	**ICC**	**(95% CI)**	**Mean difference**	**(95% CI)**	**P value [a]**
Telephone access						
Q3a How easy have you found getting through on the phone	333	0.73	(0.67, 0.78)	-2.40	(-4.91, 0.11)	0.061
Q3b How easy have you found speaking to a doctor on the phone	191	0.68	(0.59, 0.75)	-4.01	(-7.64, -0.39)	0.030
Q3c How easy have you found speaking to a nurse on the phone	82	0.63	(0.48, 0.75)	-2.85	(-8.62, 2.93)	0.330
Q3d How easy have you found getting test results on the phone	131	0.62	(0.51, 0.72)	0.25	(-3.88, 4.39)	0.903
Arriving at the appointment						
Q9 How easy do you find it to get into the building at this GP surgery or health centre?	345	0.44	(0.35, 0.52)	2.32	(0.94, 3.70)	0.001
Q10 How clean is this GP surgery or health centre?	344	0.60	(0.53, 0.66)	1.16	(-0.10, 2.42)	0.070
Q12 How helpful do you find the receptionists at this GP surgery or health centre?	335	0.69	(0.63, 0.74)	-0.60	(-2.39, 1.20)	0.514
Q13 How long after your appointment time do you normally wait to be seen?	315	0.67	(0.60, 0.73)	-0.95	(-2.60, 0.70)	0.257
Q14 How do you feel about how long you normally have to wait	308	0.70	(0.64, 0.75)	-2.11	(-4.43, 0.21)	0.074
Continuity of care						

Q 16 How often do you see the doctor you prefer	255	0.71	(0.64, 0.77)	-0.78	(-3.49, 1.92)	0.568
Opening hours						
Q18 How satisfied are you with the hours that this GP surgery or health centre is open?	325	0.65	(0.59, 0.71)	2.23	(0.40, 4.06)	0.017
Doctor-patient communication and trust						
Q22a How good was the doctor at giving you enough time	337	0.62	(0.55, 0.68)	0.45	(-0.96, 1.85)	0.532
Q22b How good was the doctor at asking about your symptoms	317	0.70	(0.64, 0.75)	-0.47	(-1.84, 0.90)	0.498
Q22c How good was the doctor at listening to you	331	0.72	(0.66, 0.77)	0.38	(-0.88, 1.63)	0.554
Q22d How good was the doctor at explaining tests and treatments	275	0.72	(0.65, 0.77)	-1.27	(-2.81, 0.26)	0.104
Q22e How good was the doctor at involving you in decisions about your care	275	0.68	(0.61, 0.73)	-1.00	(-2.65, 0.65)	0.233
Q22f How good was the doctor at treating you with care and concern	326	0.67	(0.61, 0.73)	0.23	(-1.16, 1.62)	0.745
Q22g How good was the doctor at taking your problems seriously	324	0.72	(0.67, 0.77)	-0.08	(-1.46, 1.31)	0.913
Q23 Did you have confidence and trust in doctor you saw	340	0.70	(0.64, 0.75)	-0.15	(-1.86, 1.57)	0.866
Overall satisfaction						
Q24 In general how satisfied are you with the care you get at this surgery or health centre?	344	0.74	(0.69, 0.78)	-0.58	(-1.81, 0.65)	0.353
Q25 Would you recommend this GP surgery or health centre to someone who has just moved to your local area?	333	0.77	(0.73, 0.81)	0.00	(-1.51, 1.51)	1.000

a. After applying the Holm-Bonferroni procedure with a family-wise Type I error rate of 5%, only the P value for Q9 remains significant

Discussion

Our findings show that the measurement of patient experience at practice level may mask considerable variation between doctors within the same practice. These findings are in line with other studies showing that the proportion of variance due to doctors is greater than that due to practices in the case of doctor-specific measures and less in the case of non-doctor-specific measures. [19, 215, 216] For indicators that are more likely to be under the control of the doctor (e.g. doctor-patient communication), more variance is explained by doctors than by practices: this may be taken as a validation of the use of these indicators to measure individual GP performance. Our findings additionally demonstrate that higher-performing practices usually comprise higher performing doctors, but lower performing practices may include doctors with a range of communication scores. This has important implications for evaluating practice performance, as GPs requiring support to improve their communication skills are unlikely to be identified using current practice-level approaches. As such, the current practice-based performance indicators may not provide meaningful information to commissioners, providers, or users for key domains, such as communication skills. However, other indicators observed to have more variance at practice level (e.g. cleanliness of a practice) are more suitable for evaluating performance at organisational level.

Our test-retest reliability results demonstrated good to almost perfect agreement on a number of items used in the GP Patient Survey and included within our patient experience questionnaire. Patients' willingness to recommend their practice to a friend or family member showed substantial reliability, and items orientated to staff performance also had substantial stability. Items regarding the physical environment of the practice, such as ease of access and cleanliness ranged from moderate to substantial reliability.

The results suggested that, despite the high proportion of patient-level variance in communication scores, a reliable (>0.8) adjusted mean score for individual doctors can be obtained with 46 patient scores per GP using this instrument, so little variance in reported doctor-level scores was attributable to patients and residual sources, which is in line with other published work. [207] With sample sizes smaller than this, a trade-off must be made between reliability and the utility of conducting individual- rather than group-level evaluations.

Communication is a key driver of overall patient satisfaction, [94] and ensuring patients' ability to access accurate information on performance is important if they are expected to make informed choices among providers, as current policy aspires to. In compiling performance indicators to inform patients' choice of provider, it would therefore be preferable to report communication scores at the individual practitioner level or to reliably report the range of individual practitioners' scores within an organisation. If the aspiration is to use quality indicators to identify poor performance, rather than to inform patients' choice, an alternative to the potentially costly option of obtaining communication scores for all individual practitioners could be to use organisation level assessments (such as the current GP Patient Survey) to screen for lower performing practices. Individual-level assessments could then be targeted only to organisations where performance concerns were identified. Further research to explore users', providers' and commissioners' perceptions about the feasibility of alternative approaches to generating performance data on doctor-patient communication would be useful. Furthermore, clarity about the association between the publication of performance data and quality improvement, including the mechanisms underpinning the instigation of any personal or organisational changes, is needed.

Strengths and limitations

This was a large study, including responses from 7,721 patients relating to 105 doctors across 25 practices, and producing a first report on the stability of patient responses on items used within the GP Patient Survey items over time. The stratified sampling strategy ensured participation from doctors with a range of summary scores for interpersonal skills, after adjusting for case mix, which improves generalisability to wider primary care contexts. The use of the postal survey resulted in an average delay of just over two weeks between a patient's consultation and receipt of their questionnaire, which is substantially less than the national GP Patient Survey timeframe. The two-week delay is unlikely to produce significant recall bias and would be expected to affect all participating doctors equally.

The response rate for the main study was considerably higher (51%) than achieved in the national GP Patient Survey (which ranges in recent years from 38% in 2009/10 to 35% in 2012/13), although there was substantial variation in response rates between participating GPs. The response rate for the test-retest sub-study was similar to that observed in other primary care test-retest exit surveys. [110, 207] Non-response tends to inflate doctors' and practices' scores, but this inflation is largest when non-response is highest. [217] As the lowest response rates were inclined to occur for lower-scoring practices, any non-response bias tended to attenuate the extent of variance between both doctors and practices, rather than inflate it. The estimated magnitudes of such effects were small, and it was not expected that the resulting variance at the practice and doctor levels, or for the conclusions regarding the comparison of doctor- and practice-level variances to alter.

Sampling practices from different quartiles of the GPPS practice-level communications scores may mean that the estimate of the total practice-level variance could differ slightly from that of the full population. However, this is not expected to affect the conclusions regarding the relation between practice-level scores and the extent of within-practice variation. The analysis was not adjusted for neighbourhood-level deprivation, as the research team was blinded to patients' postcodes. This limitation is unlikely to have biased the results, as deprivation has only a very small association with patients' experience after controlling for sex, age, ethnicity and health status. [209] Although the sample size of this study was considerably larger than that used for GPAS, [218] responding patients were not fully representative of the general patient population of England and Wales.

Conclusions

Currently, evaluations of GPs' communication performance most commonly report indicators at a practice level, rather than enabling patients and stakeholders to evaluate individual practitioners directly. Reporting communication-related performance indicators at practice level may mask large variation between individual practitioners. Practice-level surveys could potentially act as an initial screen for concerns about performance, with subsequent data gathering focusing on individual doctor-level surveys in lower performing practices.

Chapter 10. Exploratory trial of an intervention to improve patient experience in general practice

Abstract

Background

Our early findings and the published research evidence suggested that an intervention seeking to improve patient experience in general practice should consider the level at which feedback from patients might be provided to practice teams, whether such feedback should be facilitated, and the need for timeliness of feedback. The aim of this project was to conduct a feasibility study and an exploratory trial of an intervention which might inform change and improve patient experience in general practice.

Methods

We designed a feasibility and pilot clinical trial. Real time feedback (RTF) touch-screens were installed in practice waiting areas for 12 weeks. Practices or individual doctors received fortnightly patient feedback summaries. Some teams attended a facilitated reflection session. We undertook a multi-method evaluation of the intervention.

Results

2.5% of consulting patients provided RTF (range 0.7%-8.0% across eight practices). Men, and patients aged over-65, were under-represented among responders. Reception staff often interacted with patients but rarely encouraged touch-screen use. When staff did encourage patients to use the touch-screen, 36/60 (60%) patients attempted to start the survey. Most patients were positive about RTF but identified a range of barriers. Staff views of and engagement with RTF varied. Within-team communication influenced perceptions, and the successful implementation and use of RTF. Costs ranged from £1,125 (unfacilitated/team-level feedback) to £1,887 (facilitated/team + practitioner-level feedback).

Conclusions

Successful implementation of RTF requires team engagement, shared responsibility, and careful communication. Future studies need to make RTF accessible to a wider range of patients, and ensure that questions presented to support RTF are relevant to practices. Shorter, repeated episodes of RTF collection may be of greater utility to practices, and to researchers seeking to evaluate the approach, than sustained and ongoing RTF.

Introduction and rationale

In the original outline of this programme of research, we planned to undertake initial feasibility testing and piloting of an intervention seeking ultimately to improve patient experience of care. We anticipated that we would draw on our earlier research findings to inform the design and implementation of this project, conducted towards the end of the programme of work. In exploring what might constitute a suitable intervention, we prioritised the area of doctor communication as being one of vital importance to patients. Cheraghi-Sohi and colleagues had previously highlighted communication between doctors and their patients as being a central priority for patients in their assessment of what high quality care might look like. [54] Furthermore, our preliminary research had identified that communication, rather than access, was a key driver of patient's overall satisfaction with care. [94] Given these observations, we undertook a review of the literature seeking to identify potentially promising interventions which had targeted doctor communication as a primary consideration, and which might inform the design of an intervention study which would also incorporate findings arising from our early research from this programme.

Review of the evidence on interventions to improve communication skills in primary care

In 2008, a systematic review assessed the efficacy of feedback of real patient assessments of interpersonal care skills or brief training focused on the improvement of interpersonal care. [76] Of the nine RCTs found (two patient based feedback studies and seven brief training studies), only one feedback study (involving trainee GPs) and one training study (conducted in 1987) reported a significant positive effect. The review concluded that:

> *'the interventions to be tested in future research should consider using insights from the wider literature on communication outside primary care, might benefit from a clearer theoretical basis, and should examine the use of combined brief training and feedback to improve physicians' interpersonal skills.'*

We updated this review by repeating the search strategy in CENTRAL for 2007 onwards. We searched for studies which fitted the following criteria:

- RCTs involving primary care practitioners and their patients
- Involving one or both of the following interventions:
    a. feedback of assessments of patients on the interpersonal skills of clinicians
    b. 'brief' (up to one working week) training focussed on interpersonal care

- With a patient based assessment of change in interpersonal skills as an outcome.

Of 1610 studies returned in the search, only one study met all criteria (Table 32). [219] Haskard et. al. assessed the effect of a communication skills training programme for both patients and doctors. [219] The study involved 156 doctors from three primary care specialties (obstetrics/gynaecology, family medicine, internal medicine) in the US. Data were collected between 1996 and 1998. The clinical training programme involved three six-hour interactive workshops, conducted on a monthly basis, covering core communication skills and concepts including recognising interpersonal difficulties and tensions in doctor-patient relationships. Additionally, clinicians received three 30 to 45 minute coaching sessions involving the review of video-taped consultations. The patient training programme was a 20 minute pre-consultation intervention involving an audio CD and booklet concerning planning and organising concerns and questions to ask the doctor. There were four experimental groups:

1. Neither doctors nor patients trained (n doctors=39, control group)
2. Doctors only trained (n=41)
3. Patients only trained (n=38)
4. Doctors and patients trained (n=38)

Overall, doctor training improved doctors' information-giving and lifestyle health-behaviour counselling, and increased patients' quality of care ratings and their willingness to recommend the physician. However, doctors' satisfaction with the interpersonal aspects of their professional life *decreased* significantly more among trained versus untrained physicians. Training both doctors and patients had complex effects on doctors' satisfaction and stress: interaction effects reflected a relative increase in stress and decrease in doctor satisfaction when only one (either doctor or patient), was trained. The authors note that the intervention was intensive and may have placed additional stress on doctors, some of whom were also undergoing organisational changes at the time. [219]

Table 32. Overview of relevant studies assessing impact of interpersonal skills training

Study and setting	**Target Population and N**	**Intervention**	**Patient Satisfaction Measure** **Other Outcomes**	**Summary of findings**
Haskard et al. 2008 [219] West coast university medical centre (N=93), VA clinic (N=5), and a staff model HMO (N=58) Not stated but US (VA clinic)	Physicians from three primary care specialities (obstetrics/gynaecology, family medicine, internal medicine) N=156 Physicians; n=2196 patients in interaction	3 months of physician workshops & coaching with assessments (6hrs at each time point), with previsit intervention at preceding time 2	Patient satisfaction and perception of choice, decision-making, information, and lifestyle counselling; physicians' satisfaction and stress; global ratings of the communication process (all composite measures) No other secondary measures reported	Physician training significantly improved patients' satisfaction with information & overall care; increased willingness to recommend the physician; increased physicians' counselling (as reported by patients) about weight loss, exercise, and quitting smoking and alcohol; increased physician satisfaction with physical exam detail; increased independent ratings of physicians' sensitive, connected communication with their patients, and decreased physician satisfaction with interpersonal aspects of professional life. Patient training improved physicians' satisfaction with data collection; if only physician or patient was trained, physician stress increased and physician satisfaction decreased.
Reinders et al. 2010 [196] Vrije Universiteit (VU) University Medical Centre Amsterdam, The Netherlands	1st year General Practice Trainees (GPTs) N=53; intervention n=23 & control n=30	Patient feedback training programme (how to acquire relevant patient feedback to improve communication skills). Instructions to staff delivered in a 2 hour meeting, then half a day's instruction meeting, then 3 months to obtain 20 patient feedback.	Patient Feedback Questionnaire on Consultation Skills (PFC) Primary outcome measure MAAS-Global assessment used by trained assessors (5 behavioural scientists & 3 GPs) to assess videotaped consultations with 3x SPs (standardised patients) in 6 consultations scenarios; Process outcomes: Intensity of the GPT participation in the programme based on: the number of PFCs the GPT collected; number of learning points formulated etc. Also GPTs completed an evaluative questionnaire & National Knowledge Test in General Practice Medicine.	Consultation skills in the entire cohort of participants improved with a small-to-moderate effect size within the 3 month observation period. Consultation skills in the intervention group did not improve anymore than those in the control group. A subgroup of GPTs who participated 'actively' (i.e. intensity) in the programme showed a greater improvement in consultation skills than those who did not actively participate.

One other study assessed the effect of patient feedback on communication competencies, but used expert raters' (rather than patients') assessments of skills as the outcome (Table 32). [196] This was a trial of a patient feedback training programme in first-year GP trainees in the Netherlands. The intervention group (n=23) received instruction in how to obtain patient feedback in daily practice using the Patient Feedback Questionnaire on Consultation Skills (PFC), which focuses on GP-patient communication. Following training, GP trainees in the intervention group were asked to obtain feedback using the PFC from 20 patients over a period of three months; they also completed a self-assessment version and compared this with the patient version, and 'formulated learning points which they discussed with their GP trainers'. The control group attended the regular doctor-patient communication skills training. For the purposes of this study, simulated patients, trained to enact six consultations of moderate complexity, visited the 53 GP trainees' practices and video-taped consultations with the GP trainees. Video-taped consultations were then assessed by eight raters (5 behavioural scientists; 3 GPs) using the MAAS-Global instrument. Data on 50 GP trainees were available for analysis. Both control and intervention groups improved their consultation skills between baseline (when scores were already high) and post-intervention assessments, but there were no significant differences in improvement between control and intervention groups. However, there was a trend for intensity of participation in the patient feedback programme to predict greater improvement in MAAS-Global scores.

One of the above studies was included in a systematic review of the effect of patient feedback on physicians' consultations skills. [189] This searched for all empirical studies involving practicing doctors (including postgraduate trainees) that: incorporated feedback from real patients; assessed physicians' general consultation skills; incorporated feedback on communicative aspects in general health care, and evaluated physicians who received formal, individually directed feedback from patients (for example, by means of aggregated patient reports or educator-mediated coaching sessions).

Of fifteen studies included in the review (from 1980 to 2010), ten were in primary care, and five in other specialties. A variety of study designs were included (RCTs, quasi-experiments, cross sectional and qualitative studies). Twelve studies observed a positive effect of patient feedback on physicians’ consultation skills. In an assessment of the outcomes of studies against the Kirkpatrick hierarchy (four levels at which educational interventions can have an effect), they found that:

1. All 9 studies which evaluated Level 1 effects (valuation or views of the learning experience) reported positive effects.
2. All 4 studies which evaluated Level 2 effects (change in knowledge or skills) reported positive effects.
3. All 3 studies evaluated Level 3 effects (change in intended behaviour) reported positive effects.
4. Four of 7 studies which evaluated Level 4 effects (change in actual performance or outcomes) reported positive results.

Despite the apparently positive results, the authors argue that ‘consulting skills’ need to be much better defined in studies of this type. Additionally, observed effects cluster at the lower end of Kirkpatrick’s hierarchy, mostly in qualitative, nonrandomised studies. Actual change in performance was rarely observed. Three possible reasons were offered for the observed heterogeneity of findings:

1. Assessing actual change in general consultation skills or clinical performance may be difficult because of the lack of precision in defining ‘consulting skills’ and the lack of responsiveness of the assessment instruments.
2. Patients who have poor experiences might not report a poor outcome, limiting the effect of patient-reported outcome measures (witness the ceiling effects in many patient feedback questionnaires)
3. There may be a true absence of effect; interventions were not sufficient to drive behavioural change or doctors were not susceptible to change.

They concluded that there is a dearth of evidence showing that patient feedback has any effect on actual behaviour.

Modelling the intervention

One of the striking conclusions of our review of the previous empirical work in this area was the inconsistency in the findings of the major studies in primary care. Two of the biggest trials report opposite results, with Greco et al reporting positive effects of feedback on GP registrars, [110] and Vingerhoets et al reporting no benefits in established doctors. [192] In considering these contradictory findings, we identified two potentially important contextual factors which may moderate the relationship between intervention and outcome: (a) the training and experience of the doctors receiving the intervention (b) motivators to change (Figure 22). There is an assumption that GP communication skills can be developed at some point in the medical career, but that change becomes less likely as doctors develop a routine way of consulting, as reflected in our conceptualisation. [220]

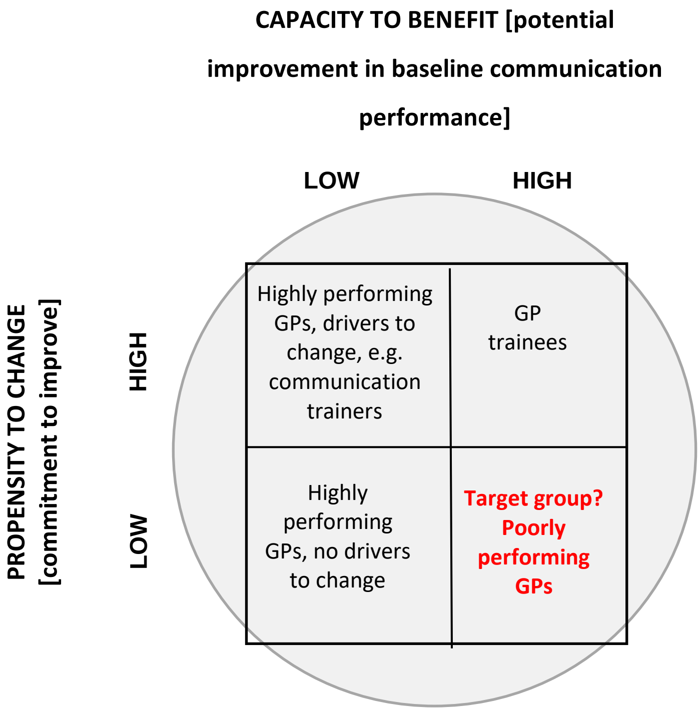


Figure 22. Typology of GPs potential responses to communication training

Discussions within the team highlighted other potential moderators, particularly the context within which any intervention might be introduced. Notably, at the time of the original programme grant application, the major contextual factor was the priority being afforded to survey results by the then Primary Care Trusts (PCTs), with the expectation that they would be engaged in project management of practices against GPPS scores. With the reorganisation of commissioning and care, this was no longer the case, and our experience with practices during the previous phases of work suggested, in the absence of external drivers, there was little support or capacity for the kind of intensive communication interventions which had been previously trialled.

We identified a number of other contextual factors that might impact on response:

- Casemix (practices serving certain patient populations or in certain areas may respond differently)
- Incentivisation of communication training and patient feedback
- Previous experience and engagement with patient feedback at a practice level
- Length of consultations and the organisation of practices in response to QoF may facilitate or hinder change
- The priority placed on access as opposed to continuity
- Practice culture and communication

Additionally, a further issue may be that educational interventions often assume that the individual practitioner is the correct ‘unit of intervention’, but it is possible that the practice may be a more important unit. However, little is known about how practices understand variation in communication quality between practitioners, or how they respond. Response may relate to complex issues around ‘sense making’, identity and the perception of legitimate work, and clinical etiquette around acceptable topics for discussion and learning among practitioners. [221, 222]

The impact of any intervention linked to communication also needed to be seen in the context of the large number of other QoF and non-QoF issues that provide competing priorities for GP time and attention.

In reflecting on the evidence to date and our own emerging findings from previous phases of work, we identified three key questions which would shape our intervention:

1. The level at which patient feedback takes place: for practices or for individual doctors
2. The requirement for feedback to be facilitated or not
3. The timeliness of feedback

We explore these in more detail below.

The level of feedback – practices or individuals?

Findings from project 2 (patient survey at practice and individual GP level: Chapter 9) identified the importance of the issue of whether any suitable potential intervention should be targeted at practice level, or at the level of the individual doctor. Through our survey, we had identified that patient feedback aggregated to practice level may mask a range of individual doctor-performance taking place within practices, especially in those practices which scored at the lower end of overall practice scores (for example in respect of patients' experiences of communication). In addition, it was unclear to us whether summarised patient feedback to clinicians was most effectively given in group, or in individual doctor settings. Given our developing interest in whether doctors or their teams were the important units for any potential intervention, we drew on our knowledge of the research literature relating to the effective functioning of teams.

One team-level attribute which can be measured and potentially manipulated to change behaviour and bring about improvements in the organisation's effectiveness is 'team climate'. [223, 224] Anderson and West [225] define team climate as "a team's shared perceptions of organisational policies, practices and procedures". Four group processes (or facets) of team climate have been proposed to be important pre-requisites for improved quality of health care: [225, 226] (i) team vision and objectives – team members' views on the clarity, sharedness and attainability and value of the team's objectives; (ii) participatory safety – team members' participation in information sharing and decision making, and psychological safety and support (for example, in trying out new ideas). (iii) task orientation – the team members' emphasis on reflection on appraisal, feedback and performance monitoring of their work, and (iv) support for innovation – perceptions of articulated and enacted support in applying new ideas and change. Studies in general practice in the UK [227, 228] and Australia [229] have reported that a favourable team climate is associated with improved standards of care for a range of long-term conditions, better access, higher patient satisfaction, higher staff satisfaction and greater perceived team effectiveness. A Finnish study of team climate in hospital settings reported that a favourable team climate was associated with lower turnover in health care staff, [230] although two more recent UK studies in general practice [231, 232] questioned the previously-observed relationship between team climate and quality of care, the authors arguing that further research focusing on the associations between team functioning and quality of care is needed.

Supported feedback to practices - Facilitated or unfacilitated?

Findings from our earlier qualitative research (see Chapter 7) identified a further important consideration – the perceived lack of support for doctors or practice teams in making sense of findings emanating from patient feedback, suggesting the potential benefit of facilitation in respect of the process of reviewing patient feedback.

Timely feedback – the potential of real time data acquisition and reporting

Collecting patient feedback is insufficient on its own to improve services, and best practice guidance [44, 85, 233] suggests that organisations need to reflect and act appropriately upon the feedback while it is still ‘fresh’. Such guidance also suggests that (i) by implementing change based on continuous real-time data, organisations can monitor whether the changes they implement have an effect on patients’ experience (ii) organisations should be prepared to show patients how their RTF has been used to change services and that (iii) sharing this information suggests that the organisation is willing to listen and respond to patient views, and that this, in turn, may mean that patients will be more willing to give their views in the future. Quite apart from the published guidance, some doctors in our early qualitative research (see Chapter 7) had identified concerns regarding the timeliness of data capture and reporting as being a potential impediment to doctors and their teams in taking action in response to patient feedback.

Recent years have seen a substantial move towards the incorporation of real-time technologies to support the acquisition of patient feedback data. Real-time feedback (RTF) involves the systematic collection, analysis and reporting of information from patients who have recently used a health care service. The approach typically uses kiosks or hand-held electronic devices (e.g. tablets) at the point of care to capture patients’ feedback about their experiences on a continuous basis. The information collected is regularly collated and reported back to the service provider to inform and support service improvement. RTF offers organisations an opportunity to improve their services by designing and delivering services to meet patients' preferences in terms of quality and content; it also enables patients to shape the services they use.

According to best practice guidance, [44, 85, 233] the collection of RTF requires careful planning, coordination and monitoring to ensure that response rates are maximised, to assess whether the patients who provide feedback are representative of the practice population as a whole, and to ensure patients are kept informed of the purpose of RTF and receive adequate practical support with the process of feeding back. One US study [234] used electronic touch screen kiosks to obtain feedback, with primary care clinic staff directing patients to the kiosk after their consultation. The approach achieved a 50% response rate and did not adversely affect waiting times or other aspect of the practice routine. Male patients were as likely as female patients to use the kiosk, but older people and ethnic minority groups were less likely to use the facility. In 2009-2010, a six-month pilot study [44] was carried out across 22 GP practices in England to determine whether real-time patient feedback could be used to help practices to understand their patients' views on services, identify opportunities to improve services, and evaluate whether any changes the practice made were effective. Three devices were piloted to collect patient feedback (tablet PC, kiosk and desktop device) and participating practices varied in size, patient list, staffing levels, geography and demography. The key findings were that RTF could be implemented successfully in most GP practices, that RTF could drive performance improvement in this setting; and that RTF has potential to complement findings from the national GP Patient Survey, but needs to be actively promoted to fully engage patients and staff.

Step-by-step guidance informed by the pilot study [44] has since been produced to provide practical advice to GP practices who wish to gather and use real-time patient feedback effectively. A number of experienced real-time technology suppliers exist in the UK, including Dr Foster's (Patient Experience Tracker or PET), The Picker Institute (Frequent Feedback service) and Customer Research Technology (ViewPoint system). Previous work highlights the need to monitor response rates and the representativeness of patients who provide RTF. For example, if hand-held devices are being handed out, some patients may be intentionally excluded, including those who are perceived as being likely to provide negative feedback or those who need extra help to feedback due to language barriers or disabilities. Some patient groups may find kiosks or hand-held devices less user-friendly and therefore decline to provide feedback if there is no assistance readily available. Others may be in a rush to leave the practice and/or reluctant to queue to leave their feedback.

Towards a clinical trial

Our early work therefore suggested testing an intervention focussing on practices in the first instance, since these were the unit of reporting for GP patient survey scores. But within that ambition, there appeared a clear need to examine issues relating to the level at which to target feedback discussions (group or individual), and to consider whether such feedback should be facilitated, or un-facilitated. Since we had originally been commissioned on the basis of undertaking a trial, internal discussions within the team identified a further important consideration – the need to consider incorporating a control group – in this case, of practices who would not, as part of our trial, receive any intervention we were testing.

Taken together, these observations suggested the potential of undertaking an intervention at practice level, using real-time data collection and feedback as the means of capturing patient feedback, and exploring within the same study the potential for group or individual feedback of results, using facilitated- or un-facilitated modes of feedback delivery. Given the timing of the research – coming at the end of seven projects, we were pragmatic in our consideration of the number practices with whom we could reasonably work. We felt that a reasonable target was to undertake feasibility and pilot testing using a randomised design in a total of 10 practices as outlined later.

Changes to study methods from original protocol

The aim of this strand of work, as stated in the original protocol, was:

“To carry out an exploratory randomised controlled trial of an intervention to improve patient experience, using tools developed in earlier parts of the programme (aim 6).”

Within the original protocol, the exact nature of the exploratory trial was therefore undefined, although we outlined three key objectives of this work stream:

1. To develop a model based on theory and published empirical evidence which relates patient assessments of interpersonal care to professional behaviours and outcomes.
2. To use that model and the views of key stakeholders from earlier workstreams to develop an intervention to improve interpersonal behaviour that is feasible and acceptable in UK primary care.
3. To conduct an exploratory trial of that intervention to:
    a. test methods for the recruitment of practices and patients
    b. test the implementation of the combined feedback and training in practice
    c. provide estimates of the effect of the intervention for sample size calculations

The exact nature of the exploratory trial we devised as a result have been broadly outline above, and are detailed below.

Aims and objectives

The aim of this workstream was to conduct, in a small number of general practices, a feasibility study and an exploratory trial of an RTF based intervention which might inform change and improve patient experience in general practice. Whilst neither phase of this work was sufficiently powered to investigate the *effectiveness* of the RTF based intervention or its various components, the workstream sought to:

1. Pilot an RTF intervention for general practices;
2. Evaluate, within the context of a pragmatic survey embedded in routine practice, and from the perspective of practice teams, the feasibility and acceptability of collecting and receiving RTF with/without a facilitated reflection session;
3. Evaluate, from the perspective of patients, the feasibility and acceptability of providing RTF, including: (a) estimating the number and proportion of patients using RTF touch-screens when these are available in practice waiting areas; (b) describing the characteristics of consulting patients who use RTF touch-screens and contrasting

these with the characteristics of all consulting patients during the same time period; and (c) obtaining the views of patients who used/did not use RTF touch-screens;

4. Estimate the costs associated with the RTF intervention from the perspectives of the NHS (cost of touch-screen equipment, training, and staff time).

The workstream had two phases:

(i) Feasibility study (January to June 2014), involving two GP practices and designed to develop the RTF intervention and research methods; and

(ii) Exploratory trial (July 2014 to February 2015), involving ten GP practices, to address the above objectives.

Methods

Practice sampling and recruitment

A similar approach for sampling practices was used in both phases of the study. Practices that fell in the lowest 50% of scores on the National GP Patient Survey (GPPS) communication items (Year 7 data) were eligible to participate. To facilitate fieldwork, practices within reasonable travelling distance of the research centres were prioritised in the initial sampling frame. In the feasibility study, two practices were purposively recruited from the South West of England to represent contrasting geographical contexts (urban/rural).

In the exploratory trial, invitations were posted to sixteen practices in the South West, and eleven practices in Cambridgeshire, and were followed up by telephone calls from the local researcher. Detailed briefing sessions were organised with staff at practices who expressed an interest. Practice managers or lead GPs provided written consent on behalf of the practice team and a Practice Profile Questionnaire was completed (providing background information about the practice).

Staff surveys

All practice staff were invited to complete a postal survey at two time points: (i) before RTF touch-screens were installed (‘baseline’); and (ii) after the 12-week RTF implementation period (‘follow-up’). Each team member was allocated a unique study ID number so that their completed questionnaires could be matched. Reminder packs were sent to non-responders approximately two weeks after delivery of the initial survey pack.

The baseline and follow-up questionnaires included demographic and contextual information (age, gender, ethnic origin, role within the practice) and the Value of Patient Feedback (VOP) Scale. The VOP scale was developed within the IMPROVE Programme specifically to measure staff attitudes towards patient feedback within this exploratory trial. The availability of a robust approach to evaluating perceptions of the utility and impact of patient feedback was central to our assessment of the engagement of health care professionals with patient experience data, and the likely impact of such information on professional practice. However, a search of the literature located no suitable approach to achieving this. We therefore drew on standard scale development processes to derive and test a new instrument. We give a brief overview of this process below – but for full details of the development of the instrument, please see Appendix 27.

Firstly, we derived key constructs using qualitative data previously collected by ourselves both within the programme and prior to the programme in other patient experience research: this gave us a body of data comprising interviews with 40 GPs and 14 focus groups with primary care practice staff concerning the impact and utility of patient experience surveys to draw on, which we supplemented with a review of relevant literature in the area. From this, we developed a pool of 56 potential items. Following expert panel review (n=6), 52 items were retained for further consideration. We undertook cognitive testing through interviews with clinicians (n=7): items were further reduced to 43, with textual amendments. Pre-testing of all 43 items took place using an online survey of doctors and nurses (n=215). Item reduction was undertaken on the basis of participant feedback and performance of the items in polychoric correlation matrices. We also undertook exploratory factor analysis resulting in further item reduction. A proposed 16 item version of the VOP scale was pilot tested in a survey of hospital doctors (n=108) and GPs and practice nurses (n=78) to inform confirmatory factor analysis. The final version of the scale used within the exploratory trial described here consisted of 16 items with five-point Likert-type rating scales (‘Strongly agree’ to ‘Strongly disagree’) (Box 8). Possible scores on the VOP range from 0 to 80.

1. Have you ever received structured patient feedback (such as through patient surveys)?
☐ At an individual level (e.g. through a report of patient feedback specific to the care you have provided)
☐ At an organisational level (e.g. through a report of patient feedback aggregated for your practice or clinic)
☐ I have never received structured patient feedback (such as through a patient survey)

2. Please put an X in one box for each row to indicate your attitude towards each statement:

	Strongly agree		**Agree**		**Neither agree nor disagree**		**Disagree**		**Strongly disagree**
1. Patient feedback is an important mechanism of quality improvement	☐	…	☐	…	☐	…	☐	…	☐
2. Making patient feedback publicly available is beneficial to other patients	☐	…	☐	…	☐	…	☐	…	☐
3. I have reservations about patient feedback received via complaints	☐	…	☐	…	☐	…	☐	…	☐
4. I have reservations about patient feedback currently received via patient forums or participant groups	☐	…	☐	…	☐	…	☐	…	☐
5. I have reservations about patient feedback currently received via surveys	☐	…	☐	…	☐	…	☐	…	☐
6. Patient surveys help identify areas for service improvement	☐	…	☐	…	☐	…	☐	…	☐
7. I can make good use of patient feedback	☐	…	☐	…	☐	…	☐	…	☐
8. Responders to patient surveys are representative of my patient population	☐	…	☐	…	☐	…	☐	…	☐
9. Feedback from current patient surveys is usually reliable	☐	…	☐	…	☐	…	☐	…	☐
10. It is beneficial to receive patient feedback via complaints	☐	…	☐	…	☐	…	☐	…	☐
11. It is beneficial to receive patient feedback via patient forums or participant groups	☐	…	☐	…	☐	…	☐	…	☐
12. It is beneficial to receive patient feedback via surveys	☐	…	☐	…	☐	…	☐	…	☐
13. I am likely to make changes to my individual practice as a result of patient feedback	☐	…	☐	…	☐	…	☐	…	☐
14. Patients are able to provide useful feedback on organisational issues, such as appointment systems	☐	…	☐	…	☐	…	☐	…	☐
15. I am concerned about my individual reputation as a result of patient feedback being made public	☐	…	☐	…	☐	…	☐	…	☐
16. Patient feedback can improve the clinical quality of care I provide	☐	…	☐	…	☐	…	☐	…	☐

Box 8. The Value of Patient Feedback Scale

Practice allocation to intervention groups

In the feasibility study, both practices piloted an intervention involving facilitated feedback and feedback reports provided at team and individual practitioner levels.

In the exploratory trial, participating practices were randomised to one of four intervention groups (8 practices) or to a control group (2 practices). The level of RTF reporting and the provision of a facilitated team reflection session (see Table 33) varied among intervention groups (A to D). Control group practices did not collect RTF during the implementation phase but could do so at the end of the project. RTF was reported at team and individual practitioner levels for control practices but no facilitated session was offered.

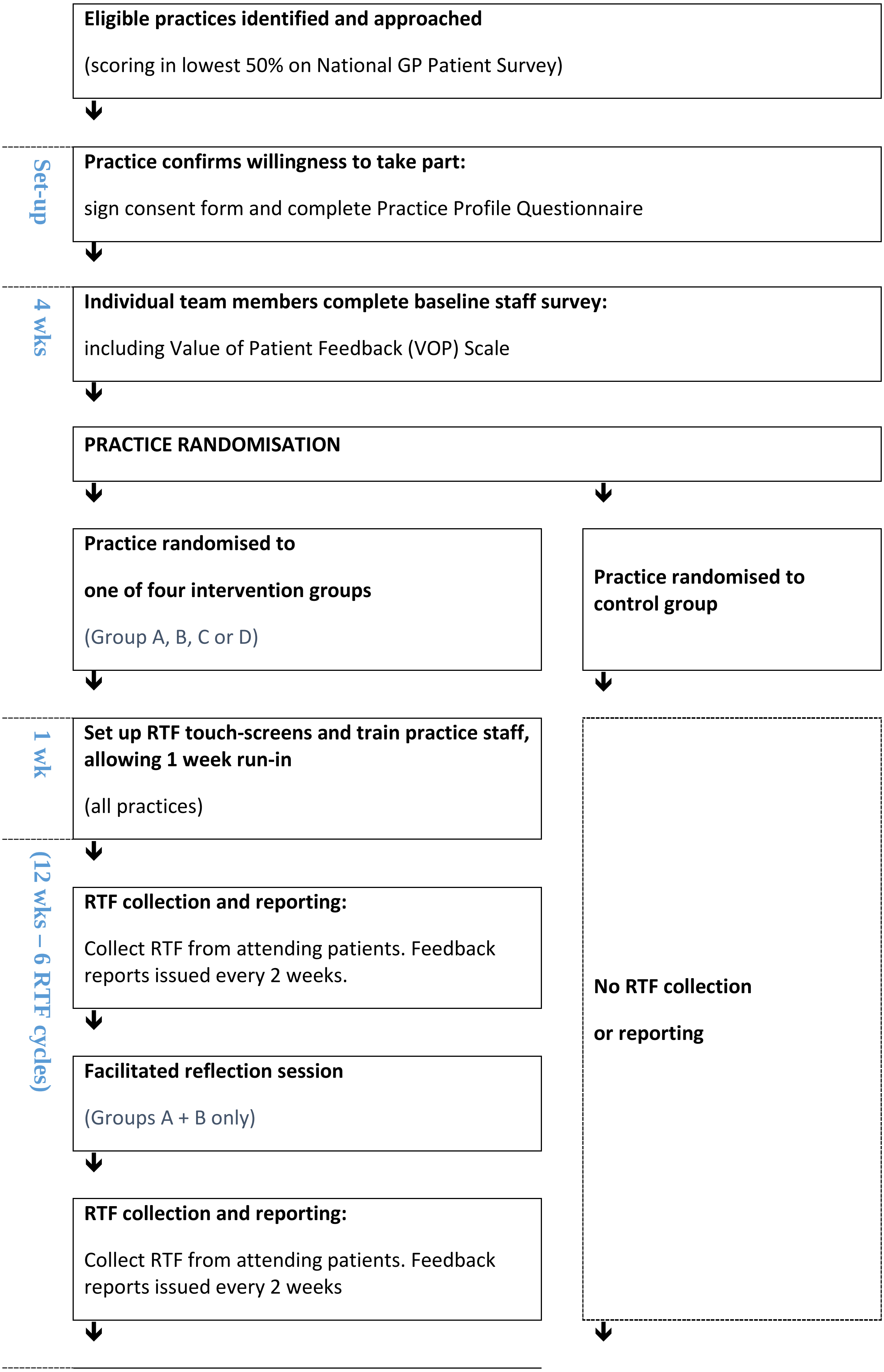
Table 33. Intervention groups in the exploratory trial phase

	Level of feedback reporting	
Facilitated reflection?	*Practice-level only*	*Practice-level plus practitioner-level*
Yes	**Group A** (2 practices)	**Group B** (2 practices)
No	**Group C** (2 practices)	**Group D** (2 practices)

After completion of the baseline staff survey, practices were randomised by a University of Exeter Medical School statistician (otherwise unconnected to the project). Randomisation occurred in two blocks of five practices using a simple randomisation approach based on random number generation. Given the small number of practices involved, stratification by variables such as practice size or GPPS score was not attempted.

After randomisation, the trial allocation was confirmed with the practice and a timeline for data collection agreed. RTF collection began at both feasibility practices in February 2014. In the exploratory trial, intervention group practices began RTF collection between July and August 2014. For the two control group practices, RTF collection began in November 2014.

Figure 23 shows the schedule of study activities at exploratory trial practices.



Eligible practices identified and approached

(scoring in lowest 50% on National GP Patient Survey)

Practice confirms willingness to take part:

sign consent form and complete Practice Profile Questionnaire

Individual team members complete baseline staff survey:

including Value of Patient Feedback (VOP) Scale

PRACTICE RANDOMISATION

Practice randomised to

one of four intervention groups

(Group A, B, C or D)

Practice randomised to control group

Set up RTF touch-screens and train practice staff, allowing 1 week run-in

(all practices)

RTF collection and reporting:

Collect RTF from attending patients. Feedback reports issued every 2 weeks.

Facilitated reflection session

(Groups A + B only)

RTF collection and reporting:

Collect RTF from attending patients. Feedback reports issued every 2 weeks

No RTF collection

or reporting

4-6 wks

Individual team members complete follow-up staff survey:

(including VOP Scale)

Focus group or interviews with practice team:

(Representation from staff in clinical and administrative roles)

END OF PRACTICE INVOLVEMENT IN EXPLORATORY TRIAL

(At the end of the study, control group practices collected RTF from attending patients over a 12-week period, with feedback reports issued every two weeks but were not offered a facilitated team reflection session)

Figure 23. Overview of practice pathway (exploratory trial)

Description of Real Time Feedback Intervention

Installation of touch-screens

In each practice, touch-screens were installed in the surgery waiting area after completion of the baseline staff survey. The installation of hardware was supported by Customer Research Technology (CRT) Limited. A short training session for practice staff was provided, explaining the purpose and day-to-day management of the touch-screens, the need to encourage patients to provide feedback, and a practical, interactive demonstration of the touch-screens. A ‘run-in’ period of up to five days allowed for any set-up issues (such as the positioning of the touch-screen) to be resolved before ‘live’ RTF collection began.

Practices were provided with leaflets and posters advertising the touch-screen to patients and were encouraged to use other means (such as the practice website or newsletter) to promote the RTF devices. Exploratory trial practices were provided with a large banner and a supply of postcards for clinical staff to hand to consulting patients.

Collection of RTF from attending patients

Patients visiting the surgery over the 12-week implementation period were eligible to provide feedback using the touch-screens, including those attending for consultations or other reasons (for example, to book an appointment). Patients activated and navigated the survey by touching the screen.

The core survey items and response options are summarised in Table 34. In line with NHS guidance, the Friends and Family Test appeared first, followed by items selected from the national GP Patient Survey. Practices could add up to two questions of their own choice. Due to limited funding, survey items were presented in English only. A parent/guardian or carer (‘proxy’) could complete the survey on behalf of the patient if necessary.

Table 34. Core RTF survey items and response options

Question Source/Type	Wording of item	Response options presented
NHS Friends and Family Test	How likely are you to recommend our GP surgery to friends and family?	Extremely likely / Likely / Neither likely nor unlikely / Unlikely/ Extremely unlikely / Don't know
GP Patient Survey (GPPS) – Telephone access	How easy is it to get through on the telephone to this practice?	Very easy / Fairly easy / Not very easy / Not at all easy /Haven't tried or Don't know
GPPS – Access to appointments	How easy is it to get an appointment for a time that suits you?	Very easy/Fairly easy/Not very easy/Not at all easy/Haven't tried or Don't know
GPPS – Receptionists	How helpful do you find the receptionists at this GP surgery or health centre?	Very helpful/Fairly helpful/Not very helpful/Not at all helpful/Don't know
GPPS – overall experience and satisfaction	Overall, how satisfied are you with the care you get at this GP surgery or health centre?	Very satisfied/Fairly satisfied/Neither satisfied nor dissatisfied/Fairly dissatisfied/Very dissatisfied
Filter question	Have you had an appointment with a health professional at the practice today?	Yes / No
Filter question	*If 'Yes':* Which of the following health professionals did you see?	Doctor / Nurse / Health care assistant / Phlebotomist (for a blood test) / Practice counsellor / Other health professional
Filter question	*If doctor or nurse:* Which doctor or nurse did you see today?	List and photographs of individual staff at the practice plus: Another doctor / Another nurse / Don't know
GPPS – confidence and trust	*If seen doctor or nurse:* Do you have confidence and trust in the doctor or nurse you saw today?	Yes, definitely / Yes, to some extent / No, not at all / Don't know or Can't say
GPPS – clinician communication skills	How good was the health professional at each of the following … (a) Giving you enough time (b) Listening to you (c) Treating you with care and concern	Very good / Good / Neither good nor poor / Poor / Very poor / Doesn't apply

	(d) Taking your problems seriously	
Practice specific items	Up to two items (with relevant response options) on topics selected by the practice team were included after the clinician communication skills items, or after the overall experience/satisfaction item (for patients who had not consulted a health professional).	
Respondent information	Are you …	The patient / Parent or guardian of the patient / Spouse or partner of the patient / Another relative or friend of the patient / Other
Patient’s gender	Are you / Is the patient …?	Male / Female
Patient’s age group	How old are you / How old is the patient?	Under 18 / 18-25 years / 26-45 years / 46-65 years / Over 65 years
Patient’s ethnic group	What is your ethnic group / What is the patient’s ethnic group?	White / Mixed / Asian or Asian British / Black or Black British / Chinese or Other
Free text comments	If you would like to leave any further comments, please type below	-

To reduce the survey length, and following discussions with our advisory group, only four of the seven GPPS communication skills items were included. Three items loading most strongly, plus one item loading least strongly, onto overall communication scores for GPs and nurses were selected. [119] Filter questions were included to ensure respondents were presented with items relevant to their visit. For example, patients who had not had a consultation were not asked to rate the communication skills of a health professional.

Practice feedback reports

All practices received a fortnightly summary of team-level feedback (six reports per practice in total). Patient feedback was transmitted from the touch-screens to CRT Limited via Wi-Fi or G3 connections. Where no reliable signal was available, data were manually downloaded (approximately fortnightly) by the researcher.

Data were ‘quarantined’ if (for example) the respondent had not answered a minimum number of survey items, if it response options appeared to be randomly selected, or if a response had been provided in a time frame that suggested the question could not have been read. Otherwise data was considered to be ‘valid’.

Cumulative feedback reports were generated by CRT Limited including all valid feedback collected since touch-screen installation. Reports contained frequency tables and graphs, and patients’ free-text comments. Free-text comments were screened by the local researcher and details that might identify individual patients removed. Negative comments about a clinician’s practice or standards of care were discussed on an individual basis with the Chief Investigator and a course of action proportionate to the risk to patients was agreed.

GPs, nurses and health care assistants (HCAs) from the two feasibility and six exploratory trial practices (Intervention groups B and D, or Control group) were provided with personalised reports if they accumulated valid feedback from 20 or more respondents. These were similar in format to the team-level reports, but summarised feedback only from patients who reported consulting the practitioner.

Team-level reports were e-mailed to the Practice Manager every fortnight for dissemination to the wider practice team. Personalised reports were e-mailed or posted direct to the individual practitioner.

Facilitated team reflection session

Two feasibility and four exploratory trial practices (intervention groups A and B) were offered a facilitated team reflection session. Facilitated reflection sessions took place at the surgery approximately half way through the RTF implementation (Weeks 6-7), and lasted 45-60 minutes. Clinical and administrative staff were invited to attend, and participants were provided with printed copies of the practice's most recent team-level feedback report. The session was led by one of four experienced GP appraisers/trainers based in Exeter or Cambridge. Facilitators were briefed in advance about the study and the aims of the session, and were provided with information about the practice and their most recent RTF report. The facilitator and practice team explored the feedback and identified aspects of service which were well-received by patients, as well as areas with potential for improvement. Following experience in the two feasibility study practices, structured action-planning paperwork was used in the facilitated session for the exploratory trial. One member of staff was nominated to complete an Action Plan sheet during the session, summarising the team's reflections, discussions and agreed action points. With the practice team's permission, the session was observed by a researcher who took brief field notes.

Details of data collection

A multi-method approach (see Figure 24) was adopted to investigate the feasibility and acceptability of the RTF interventions. This included focussed ethnographic methods to explore how the new technology can be introduced into a complex system with 'multiple human actors'. [235]

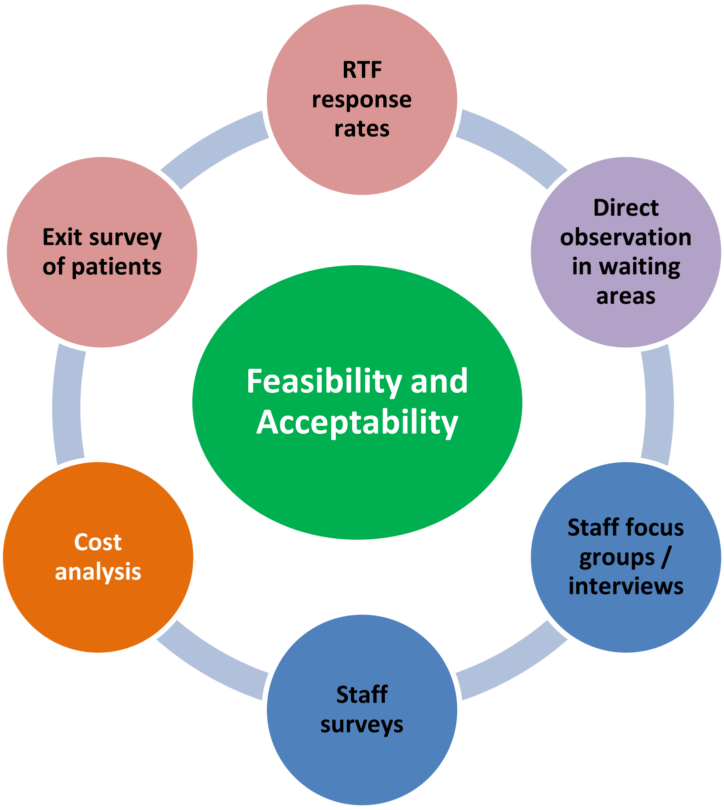


Figure 24. Multi-method approach to data collection

Practice visits

During the 12-week implementation period, researchers visited participating practices every fortnight to observe patients' interactions with practice staff and RTF devices. The visits took place on varying weekdays and at a range of times, to ensure different staff were on duty, and to capture workload variations and a range of activities at the practice. All data were recorded in anonymised form to protect patient and staff confidentiality. The practice visits were divided into shorter sessions (approximately one hour), each focusing on different types of data collection.

Unstructured observations

Researchers took detailed, contemporaneous field notes describing the practice environment, as well as interactions between patients and staff, and patients and RTF devices. These included descriptions of specific events, as well as the researcher's own impressions and interpretations. Observation notes were periodically shared and discussed within the research team to develop the methodology and maximise the richness of data collected.

Structured observations

Researchers used checklists, including Yes/No tick boxes [236] to systematically record interactions between patients and practice staff, and patients' use of the touch-screens and publicity materials (Box 9).

(i) Patient interacts with a receptionist during their visit (for any reason)
Where some interaction occurred, did the receptionist:
  - Tell the patient about the opportunity to leave RTF?
  - Point to or take the patient to the touch-screen?
  - Offer to demonstrate how to use the touch-screen?

(ii) Patient interacts with a health professional in the waiting area (for any reason)
Where some interaction occurred, did the health professional:
  - Tell the patient about the opportunity to leave RTF?
  - Point to or take the patient to the touch-screen?
  - Offer to demonstrate how to use the touch-screen?

(iii) Patient interacts with RTF publicity materials
During their visit, did the patient:
  - Pick up a RTF leaflet/flyer or look at an RTF poster
  - Spend time reading RTF information in detail

(iv) Patient interacts with touch-screen equipment
Was a touch-screen free when the patient left the surgery?
What level of interaction was observed:
  - Patient looked at or walked up to a touch-screen
  - Patient stopped to read the first screen
  - Patient touched the first screen to begin the survey
  - Patient stopped using the touch-screen without answering any items
  - Patient answered some or all RTF items

(v) Level of assistance required with touch-screens
  - Patient asked (staff or researcher) what the touch-screen was for
  - Patient required help to use the touch-screen (from staff or researcher)

Box 9. Interactions and events targeted during structured observation sessions

During observation sessions, a poster was displayed in the waiting area to explain the researcher's presence in the practice. Individual patient and staff consent was not sought, in case this significantly altered behaviour relating to the touch-screens. [237]

Patient exit surveys

Researchers conducted brief face-to-face exit surveys with a convenience sample of patients as they left the practice, whether or not they had used the RTF device. The purpose of the survey was explained to patients and their verbal consent to participate was sought. Participants responded verbally to a series of structured questions about the touch-screens and their views of RTF (see Table 35). Brief demographic details were also recorded.

Table 35. Summary of patient exit survey items

	Attending patients who:	
	Had used a touch screen	**Had not used a touch screen**
Was the patient aware of the opportunity to leave feedback using a touch-screen device?	N/A	✓
How did the patient find out about the opportunity to leave feedback?	✓	✓
Reasons for not using the touch-screen today	N/A	✓
How easy did the patient find it to use the touch-screen?	✓	N/A
Did the patient have any difficulty understanding the RTF questions?	✓	N/A
How long did it take to answer the RTF questions?	✓	N/A
The patient's overall view of touch-screens as a way of collecting patient feedback	✓	✓
Patient's gender	✓	✓
Patient's age group	✓	✓

Extraction of appointment statistics

Researchers used practice appointments systems to determine the number of appointments attended (with any health professional) during the 12-week RTF implementation period, and to collect anonymised age/gender information about the patients who had consulted. This information was used, combined with consultation information from the RTF survey, to calculate the percentage of consulting patients who had used a touch-screen, and to explore the extent to which those patients were representative of the consulting patient population, in respect of age and gender. No records were kept of the number of patients who attended the practice for other reasons so it was not possible to calculate a ‘feedback rate’ for this group.

Practice team focus groups or interviews

At the end of RTF implementation, researchers conducted either semi-structured interviews with a purposive sample recruited from all practice staff, or focus groups to which all practice staff were invited. Interviews and focus groups explored aspects of the RTF implementation within the practices, including training and technical support, processes involved in the collection of RTF, reports from the devices, learning from and acting on patient RTF, and how much staff valued the feedback they received. Interviews and focus groups lasted approximately 40-45 minutes, were audio-recorded with the participants’ permission and transcribed verbatim. Participants provided individual written informed consent prior to the focus group or interview. Lessons from discussions with the feasibility practice teams were used to refine the RTF intervention and wider study processes prior to the exploratory trial phase.

Interviews with facilitators

Semi-structured interviews were conducted with facilitators to explore aspects of the team reflection sessions, including: the facilitator’s general approach to the session; perceptions about practice teams’ engagement with the facilitation process; views about the practice team’s or individuals’ understanding of and reflections about RTF (including individualised feedback) and any plans the team had for acting on it; and their assessment of the value of facilitated sessions in general. Facilitators provided individual written informed consent, the discussion was audio-recorded with their permission, and transcribed verbatim.

Data analysis

Quantitative analysis

The proportion of consulting patients who used the touch-screens during the RTF implementation period was calculated (overall and for each practice) using the following equation:

$$\frac{\textit{Number of patients who provided valid RTF } \underline{\textit{and}} \textit{ reported having a consultation with a health professional (ascertained from each practice's final RTF dataset)}}{\textit{Number of patients who consulted a health professional in the same period (ascertained from each practice's computerised appointments system)}}$$

The age, gender and ethnic origin of consulting patients who provided valid feedback was ascertained from each practice's final RTF report, along with the type of health professional (GP/nurse/HCA/other) consulted.

To identify whether particular patient groups were more likely to use the touch-screens than others, the proportions of patients who provided valid RTF over age groups and genders was compared with the respective proportions of all patients who consulted in the same time period (ascertained from the appointments statistics) using z-tests.

Not all patients who provided RTF disclosed their age or gender. To derive more accurate response rates and proportions of responders by gender and age bands, the number of responders in each demographic sub-group per practice was increased in proportion to the number of missing values expected in that sub-group, based on the proportions in the practice's consulting population. For example, if a practice had ten respondents who had not provided their gender, and 60% of appointments at the practice were for females, the number of responding females for that practice was increased by six, and the number of responding males was increased by four.

Data derived from completed structured observation checklists were summarised descriptively to determine the frequency of a range of pre-specified interactions occurring during the observation periods – for example, the number (%) of patients who were encouraged by reception staff to use the touch-screen. Patient exit survey responses were also summarised descriptively: the number (%) percentage of patients endorsing each response option.

To determine whether the RTF intervention was associated with staff attitudes to patient feedback, mean scores on the VOP scale before and after the study were compared using a paired-samples *t*-test.

We explored whether changes over time in VOP scores varied across trial arms and health professional groups using analysis of variance (ANOVA). The ANOVA model included change in VOP score as the dependent variable, derived by subtracting baseline scores from follow-up scores, and therefore included responses only from staff who returned baseline and follow-up questionnaires. The model had a 5 (Groups A, B, C, D, and control) x 2 (clinical, non-clinical) design, with trial arm and practice role as between-subjects factors. To provide more balanced 'practice role' groups, GPs, nurses and other health professionals were categorised as having a clinical role, while receptionists, administrators and managers were categorised as having a non-clinical role.

Cost analysis

This analysis sought to estimate the cost of providing an RTF intervention in GP practices over a 12-week period. Such costs could potentially be compared with outcomes in a cost-consequences analysis. Cost items are listed in Table 36.

Table 36. Cost analysis – cost items

Provision and hire of equipment & feedback	Training/set up	Facilitated reflection
Publicity (leaflets and poster)	Time for:	Facilitator fee
Kiosk rental	GP	Staff time (categories as per training/set up)
Touchscreen rental	Practice manager	
Kiosk collection	Practice nurse	
Reporting	Receptionist	
	Healthcare Assistant	
	Administrator	

Data for the hire of equipment and provision of team- and individual-level reports were provided in aggregate by the RTF provider. Time inputs for practice staff and facilitators were collected from each of the eight intervention practices. Unit costs for staff (Table 37) were extracted from standard UK sources. [238] The price year for the analysis was 2014 and costs included VAT where applicable. The cost in the two control practices was assumed zero. Given the pilot nature of the study and the small sample size, summary costs only were reported and no attempt was made to draw comparisons between trial arms.

Table 37. Cost analysis – unit costs

Role	**Hourly rate**	**Source / notes**
GP	£109.00	PSSRU 2014* p195, per hour of GMS activity, excluding direct care staff costs and qualification costs.
Nurse practitioner	£51.00	PSSRU 2014* p193. Nurse advanced per hour (excluding qualification costs).
Practice Nurse	£34.00	PSSRU 2014* p192. Per hour (excluding qualification costs).
Care Assistant	£10.06	PSSRU 2014* p266, mean pay for health care assistants: £16,600, assume 37.5hours per week and 44 weeks per annum.

Physiotherapist	£32.00	PSSRU 2014* p179, per hour
Pharmacist	£51.00	PSSRU 2014* P184, per hour
Practice Manager	£21.54	Mid-point (point 30) AfC Band 7** PSSRU 2014* p197 states practice manager for a dentist typically AFC Band 7. Assumed same cost for a GP practice and 37.5 hours per week, 44 weeks per annum.
Administrator	£10.78	Assumed midpoint AFC band 3 (£17794), 37.5 hours per week and 44 weeks per annum**
Receptionist	£9.35	Mid-point AfC Band 2 (£15432). PSSRU 2014* p197 states receptionist for a dentist typically AFC Band 2. Assume same for GP practice and 37.5 hours per week, 44 weeks per annum.

* Curtis 2015 [228] ** Royal College of Nursing

Qualitative analysis

NVivo v.10 software (QSR International Ltd, 2012) was used to facilitate the organisation, coding, linking and retrieval of the qualitative data from the sources described above. After initial independent reading of a sample of the transcripts, two qualitative researchers (MC and AD) discussed preliminary themes, eliminated any duplication and resolved any differences. MC developed a coding framework, underpinned by Normalisation Process Theory (NPT) constructs [239] into which the refined themes were organised. Using this framework, the remaining transcripts were each analysed and coded by MC, and a subset by AD. Both researchers ensured that data which did not appear to fit within the NPT constructs were also included in their analysis. Progress of the analysis was discussed in regular group sessions with a third researcher (CW), whose role as academic lead afforded a comprehensive view of qualitative and quantitative aspects of the RTF research.

Normalisation Process Theory (NPT) provided an analytical tool with which the RTF implementation was explored. NPT is a theory of implementation originally developed to understand the embedding of new technologies into health systems, [239] so was judged to be particularly appropriate to the processes involved in the RTF implementation. The theoretical framework includes four constructs: coherence/sense making, cognitive participation/relationships, collective action/enacting, reflexive monitoring/appraisal (Figure 25). Although presented as discrete, linear categories, in reality the NPT constructs often operate and are experienced simultaneously.

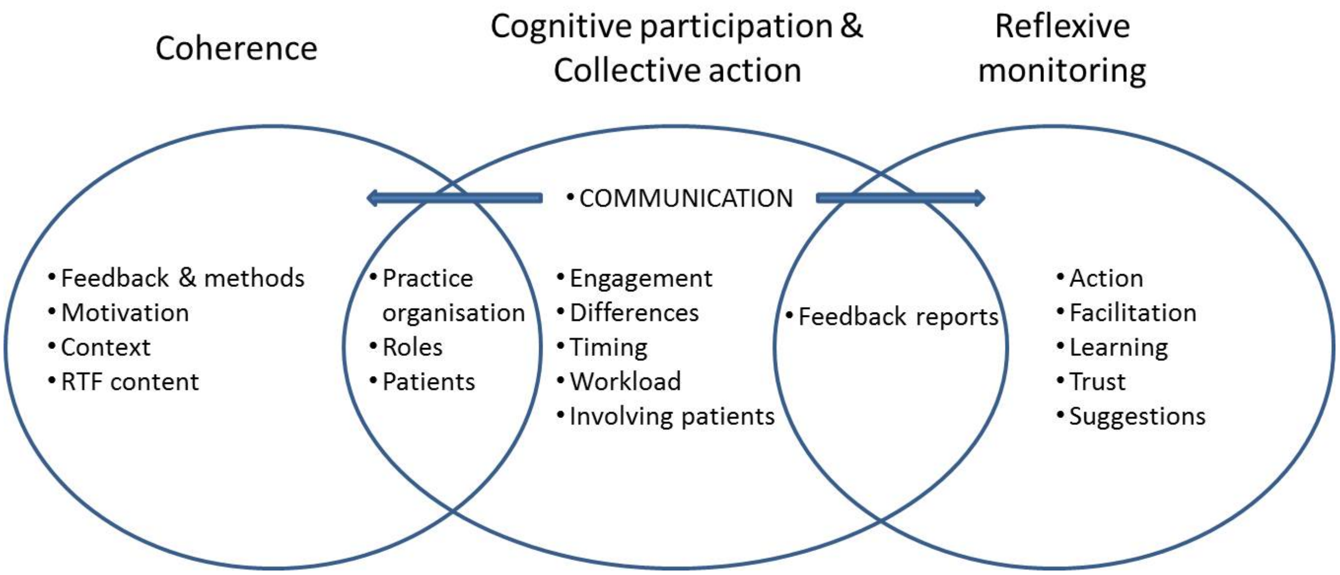


Figure 25. Normalisation Process Theory framework – qualitative analysis

Results: feasibility study

Summary of findings

The characteristics of the two practices recruited for the feasibility study are summarised in Table 38. Across the two practices, 607/14,372 (4.2%) consulting patients provided valid feedback and the rate of touch-screen use in each practice was similar – 178/4,355 (4.1%) in Practice A and 429/10,017 (4.3%) in Practice B.

Table 38. Characteristics of participating practices (feasibility phase)

	Practice A	**Practice B**
List size*	4,122	6,555
Number of practice staff:		
GPs	5	9
Nurses	1	1
HCAs	3	4
Reception/administrative	11	22
Managerial	1	5
Setting	Urban	Rural
GPPS centile score	46.7%	16.7%
Deprivation decile**	7	6
Proportion of telephone consultations	76-100%	26-50%

*Average list size for England is 7,041.

** Data from 'National General Practice Profiles' (Public Health England); lower numbers indicate more deprivation.

Observation of interactions in the reception/waiting areas revealed that, whilst staff interacted with patients (100/185; 54% observations), they were rarely encouraged to leave feedback or directed to the touch-screens (4/87; 5% interactions with reception staff). RTF publicity materials were rarely noticed by patients (2/185; 1% observations) but were competing with a large volume of health and social care information displayed in waiting areas. This suggested that more conspicuous publicity materials and a greater emphasis in set-up training on the need to encourage patients to use touch-screens were required. Large pull-up banners and postcards for clinicians to hand to consulting patients were therefore introduced for the exploratory trial phase.

Patients were divided in their views of RTF as a way to providing feedback. Many were positive about the touch-screens (51/60; 85%), finding their immediacy and anonymity advantageous, and welcomed the opportunity to provide feedback. However, some patients highlighted potential problems for other patients who may not be comfortable with computers, who are rushed, or those with consecutive appointments. Patients who did not use the touch-screens commented on the positioning of the devices and feeling "like you are on show". These comments were taken into consideration when advising practices in the exploratory trial phase.

Staff from both practices were enthusiastic about the touch-screens, but confirmed that the publicity materials did not adequately attract patients' attention. Receptionists found juggling their normal workload with encouraging patients to use the touch-screens difficult. Although staff found the facilitation sessions useful, circulation of RTF reports to the wider team and identifying an action plan at the end of the session were problematic. It was suggested, for the exploratory trial, that one individual within each practice should be responsible for completing an action plan sheet for circulation to the team after the facilitation session.

Results: exploratory trial

Ten practices were recruited for the exploratory trial: eight from the South West and two from Cambridgeshire. Table 39 summarises the characteristics of participating practices.

Table 39. Characteristics of participating practices (by exploratory trial group)

	Intervention A		**Intervention B**		**Intervention C**		**Intervention D**		**Control Group**	
	Practice 1	Practice 2	Practice 3	Practice 4	Practice 5	Practice 6	Practice 7	Practice 8	Practice 9	Practice 10
List size*	4,114	4,568	3,618	8,005	13,000	15,189	10,998	9,500	11,727	6,675
Number of practice staff:										
GPs	3	4	3	6	11	6	12	6	6	4
Nurses	2	3	1	5	3	7	7	3	8	2
Health care assistants	2	1	1	1	3	2	2	2	2	1
Reception/administrative	6	8	7	12	12	12	17	14	16	8
Managerial	1	3	2	2	3	2	3	2	2	1
Setting	Rural	Urban	Urban	Inner city	Rural	Urban	Urban/ rural	Inner city	Urban	Urban
GPPS centile score	33.4%	39.5%	28.9%	14.3%	34.0%	31.9%	21.9%	32.7%	27.1%	14.5%
Deprivation decile**	8	2	10	2	6	2	9	7	7	7
Consultations per week, mean	441.6	707.0	181.3	620.3	1,809.5	474.6	434.0	636.8	1,040.2	250.3

*Average list size for England is 7,041. ** Data from 'National General Practice Profiles' (Public Health England); lower numbers indicate more deprivation.
Practice 8 did not provide numbers of staff employed.Intervention Group A = facilitated reflection, practice-level feedback; Group B = facilitated reflection, practice-level and individual practitioner level feedback; Group C = unfacilitated reflection, practice-level feedback; Group D = unfacilitated reflection, practice-level and individual practitioner level feedback.

Proportion and characteristics of patients providing Real Time Feedback

Altogether, 1,941/79,145 (2.5%) consulting patients provided valid feedback (‘responders’), with a 95% confidence interval (CI) of 2.3%-2.6%. Patient use of the touch-screens varied across practices (Table 40), with a range of 0.7% (95% CI: 0.6%-0.9%) to 8.0% (95% CI: 7.3%-8.8%). The mean practice-level response rate was 3.2% (s.d. = 2.2).

Data on patient ethnicity were not available from the appointments system at any of the practices and, at three practices, appointments data could not be broken down by age and gender. Table 40 shows the response rate broken down by gender and age bands, using data from seven of the ten exploratory trial practices. For these practices, the mean percentage of responders who did not provide their gender was 6.7% (range 1.9%-13.7%), and 6.7% (range 2.2%-13.7%) did not provide their age.

Table 40. Real-time feedback completion rates for consulting patients (exploratory trial)

	RTF responses / Appointments	% (95% CI)
Overall	1,941 / 79,145	2.5 (2.3-2.6)
Practice (group)		
1 (Intervention A)	231 / 5,299	4.4 (3.8-4.9)
2 (Intervention A)	201 / 8,484	2.4 (2.1-2.7)
3 (Intervention B)	110 / 2,175	5.1 (4.2-6.1)
4 (Intervention B)	168 / 7,443	2.3 (1.9-2.6)
5 (Intervention C)	162 / 21,764	0.7 (0.6-0.9)
6 (Intervention C)	64 / 5,695	1.1 (0.9-1.4)
7 (Intervention D)	416 / 5,208	8.0 (7.0-8.8)
8 (Intervention D)	102 / 7,642	1.3 (1.1-1.6)
9 (Control)	386 / 12,482	3.1 (2.8-3.4)
10 (Control)	101 / 3,003	3.3 (2.7-4.1)
Gender[1]		
Men	531 / 23,739	2.2 (2.1-2.4)
Women	859 / 34,226	2.5 (2.3-2.7)
Age band[1]		
Under 18 years	150 / 6,747	2.2 (1.9-2.6)
18-25 years	78 / 3,998	2.0 (1.5-2.4)

26-45 years	315 / 12,383	2.5 (2.3-2.8)
46-65 years	469 / 15,190	3.1 (2.8-3.4)
Over 65 years	377 / 19,647	1.9 (1.7-2.1)

[1] Appointments data could not be broken down by gender or age for 3 out of 10 pilot practices. As such, the numbers displayed do not sum to the overall totals given.

Intervention Group A = facilitated reflection, practice-level feedback; Group B = facilitated reflection, practice-level and individual practitioner level feedback; Group C = unfacilitated reflection, practice-level feedback; Group D = unfacilitated reflection, practice-level and individual practitioner level feedback

The age and gender of consulting patients who provided RTF (at seven of the ten practices) are summarised in Table 41, together with the characteristics of all patients from these practices who consulted during the study period.

Table 41. Representativeness of consulting patients who provide real-time feedback

Characteristic, n (%)	**Responders out of total (%)**	**Proportion in population (%)**	***P*-value***
Women**	859/1,390 (61.8)	34,226/57,965 (59.0)	**0.039**
Age band**			
Under 18 years	150/1,390 (10.8)	6,747/57,965 (11.6)	0.329
18-25 years	78/1,390 (5.6)	3,998/57,965 (6.9)	0.061
26-45 years	315/1,390 (22.7)	12,383/57,965 (21.4)	0.243
46-65 years	469/1,390 (33.7)	15,190/57,965 (26.2)	**<0.001**
Over 65 years	377/1,390 (27.1)	19,647/57,965 (33.8)	**<0.001**
Ethnicity***			
White	1,724/1,941 (88.8)	n/a	-
Mixed	28/1,941 (1.4)	n/a	-
Asian	52/1,941 (2.7)	n/a	-
Black	27/1,941 (1.4)	n/a	-
Chinese	8/1,941 (0.4)	n/a	-
Missing	102/1,941 (5.2)	n/a	-

* Proportions (z) test

** Appointments data could not be broken down by gender or age for 3/10 exploratory trial practices. These proportions are taken from the real-time feedback and appointments data of the 7 remaining practices.

*** Appointments data could not be broken down by ethnicity for any practice.

There was a higher proportion of female responders (61.8%) than in the consulting population (59.0%), *z*=2.063, *p*=0.039. The proportion of responders in the under 18, 18-25, and 26-45 age bands did not differ significantly from the proportions in the consulting population. There were significantly more responders aged 46-65 years (33.7% of responders compared to 26.2% of the population; z=6.300, *p*<0.001), and significantly fewer responders aged 65 years or more (27.1% of responders, compare to 33.8% of the population; *z*=-5.277, *p*<0.001).

Observed patient and staff interactions

Researchers conducted structured observation sessions only at the eight intervention group practices in the exploratory trial. Observations were not conducted at the control group practices.

In total, 873 of 1,205 (72.5%) attending patients were observed to have some form of verbal interaction with a receptionist, but fewer interactions with health professionals in the waiting area (0.8%). Across 1,199 observed staff-patient interactions, 60 (5%) patients were encouraged to use the touch-screens by a receptionist, but never by a health professional. When staff encouraged patients to use the touch-screen, 36/60 (60%) patients attempted to start the survey. In contrast, only 28/1,114 (2.5%) patients attempted the survey without encouragement. Few patients (78/1,199; 6.5%) were observed to read publicity materials in the waiting area.

Patient views of RTF

In total, 375 patients participated in exit surveys at the eight intervention arm practices in the exploratory trial. Of those surveyed, 103 (27.5%) had used the touch-screen in the waiting area and 272 (72.5%) had not.

Of the patients who had used a touch-screen, 87/101 (86.1%) had positive views of RTF as a way of leaving feedback for the practice. All responders reported they had found it easy to complete the RTF survey and that they answered all questions. The majority (79/98; 80.6%) of responders reported completing the survey in two minutes or less.

Patients who had not used a touch-screen gave a range of reasons for this. Over half (149/268; 55.6%) were not aware of the touch-screens or the opportunity to leave feedback. Of those who were aware of the touch-screens, 29/84 (34.5%) said they did not have time to use them; 5/84 (6.0%) felt their feedback would not be relevant (for example, because it was positive); 4/84 (4.8%) had concerns about anonymity or how the feedback would be used; 15/84 (17.9%) had concerns about technology; and 12/84 (14.3%) reported completing RTF before but were not aware they could leave feedback on each visit. Despite not using the touch-screens during their current visit, 178/260 (68.5%) patients thought the idea of RTF was good.

RTF and staff attitudes to patient feedback

Across the ten exploratory trial practices, 162/247 (65.6%) members of staff returned a baseline questionnaire, 123/247 (49.8%) returned a follow-up questionnaire, and 107/247 (43.3%) returned both questionnaires. Of these, 92/107 (86.0%) completed all items on the VOP scale at both time points and were included in the analysis of pre- and post-intervention scores. Table 42 presents mean VOP scores at the two time points for each practice and by staff group. The results suggest that staff perceptions of the value of patient feedback did not change significantly from baseline (mean=42.9, s.d.=8.44) to follow-up (mean=41.7, s.d.=8.20), t_{91}=1.703, p=0.092.

Table 42. Value of patient feedback scores for practice staff: pre- and post-intervention, broken down by trial arm and staff group

	n	Pre-intervention Mean (s.d.)	Post-intervention Mean (s.d.)	Difference Mean (s.d.)	95% CI for the difference
Group A					
Clinical	4	45.25 (13.6)	41.00 (14.02)	-4.25 (11.03)	-21.8, 13.30
Non-clinical	8	40.25 (6.18)	41.63 (5.85)	1.38 (2.88)	-1.03, 3.78
Overall	12	41.92 (8.99)	41.42 (8.69)	-0.50 (6.79)	-4.81, 3.81
Group B					
Clinical	6	41.17 (8.28)	44.50 (9.27)	3.33 (10.27)	-7.44, 14.11
Non-clinical	6	45.17 (3.31)	39.33 (3.98)	-5.83 (6.62)	-12.78, 1.11
Overall	12	43.17 (6.37)	41.92 (7.32)	-1.25 (9.53)	-7.30, 4.80
Group C					
Clinical	14	42.29 (10.23)	40.86 (11.44)	-1.43 (4.47)	-4.01, 1.15
Non-clinical	8	42.63 (6.09)	40.75 (4.37)	-1.88 (4.26)	-5.43, 1.68
Overall	22	42.41 (8.79)	40.82 (9.35)	-1.59 (4.29)	-3.50, 0.31

Group D					
Clinical	5	45.80 (8.20)	45.80 (9.04)	0.00 (5.24)	-6.51, 6.51
Non-clinical	12	43.50 (5.60)	43.50 (7.24)	0.00 (6.4)	-4.06, 4.06
Overall	17	44.18 (6.29)	44.18 (7.59)	0.00 (5.92)	-3.04, 3.04
Controls					
Clinical	16	40.94 (11.89)	40.63 (9.08)	-0.31 (8.15)	-4.65, 4.03
Non-clinical	13	45.15 (7.10)	41.62 (6.84)	-3.54 (5.33)	-6.76, -0.32
Overall	29	42.83 (10.09)	41.07 (8.03)	-1.76 (7.10)	-4.46, 0.94

“Clinical” grouped GPs, nurses and other health professionals; “non-clinical” grouped receptionists, administrators and managers.

Group A = facilitated reflection, practice-level feedback; Group B = facilitated reflection, practice-level and individual practitioner level feedback; Group C = unfacilitated reflection, practice-level feedback; Group D = unfacilitated reflection, practice-level and individual practitioner level feedback

The ANOVA examining the difference between follow-up and baseline VOP scores assessed change over time between trial arms and staff groups. There was no significant effect of trial arm, $F_{4,77}=0.301$, $p=0.877$, indicating that change over time in VOP scores did not vary reliably across the intervention and control groups. Similarly, there was no significant effect of staff role, $F_{2,77}=2.351$, $p=0.102$, suggesting that neither the VOP scores of clinical staff nor those of non-clinical staff changed over time. No significant interaction between trial arm and staff group was apparent, $F_{8,77}=1.549$, $p=0.154$.

Cost analysis

Costs by RTF intervention groups A to D are shown in Table 43. The mean cost per practice of providing RTF was approximately £1,117 over the 12-week intervention period. The largest component was rental of the RTF touch-screens (total £972 per practice). The practice manager and administrative staff attended the set-up session in most practices. At practices allocated to a facilitated feedback arm (Groups A and B), GPs and nurses also attended (see Figure 26). A nurse was the only attendee at the set-up session in Practice 6, and this practice had the lowest feedback response rate by the end of RTF implementation. Training time was assumed at 15 minutes per staff member, estimated at £27 per practice, SD £22.

Table 43. Cost analysis – results

	Group:	**A**	**B**	**C**	**D**	**All groups** mean (SD)	Groups
	Feedback level*:	P	P&I	P	P&I		
	Facilitated session?	Yes	Yes	No	No		
Item	n practices	2	2	2	2	8	
RTF equipment – hire and provision							
Publicity (posters & leaflets)	750 postcards + 1 poster per practice	£107					
Touch-screen (kiosk) rental	12 week hire	£630					
Touch-screen (desk-top) rental	12 week hire	£342					
Kiosk collection	-	£38					
Reporting	-	£75**					
Total						£1,117	(A – D)
Practice staff set-up session		£43	£34	£8	£22	£27 (£22)	(A - D)
Total						**£1,144 (£22)**	**(A - D)**
Facilitated reflection							
Facilitator fees		£250	£250			£250 (£58)	(A, B)
Practice staff to attend facilitation		£477	£378			£428 (£180)	(A, B)
Total		£727	£628			**£678 (£227)**	**(A, B)**
Total cost		**£1,887**	**£1,779**	**£1,125**	**£1,139**		

**P: Practice level reports provided; P&I: Practice- and individual-level reports provided. ** Cost of reporting was averaged over all eight intervention practices. The marginal cost of individual level feedback over group level feedback was assumed to be zero.*

Group A = facilitated reflection, practice-level feedback; Group B = facilitated reflection, practice-level and individual practitioner level feedback; Group C = unfacilitated reflection, practice-level feedback; Group D = unfacilitated reflection, practice-level and individual practitioner level feedback

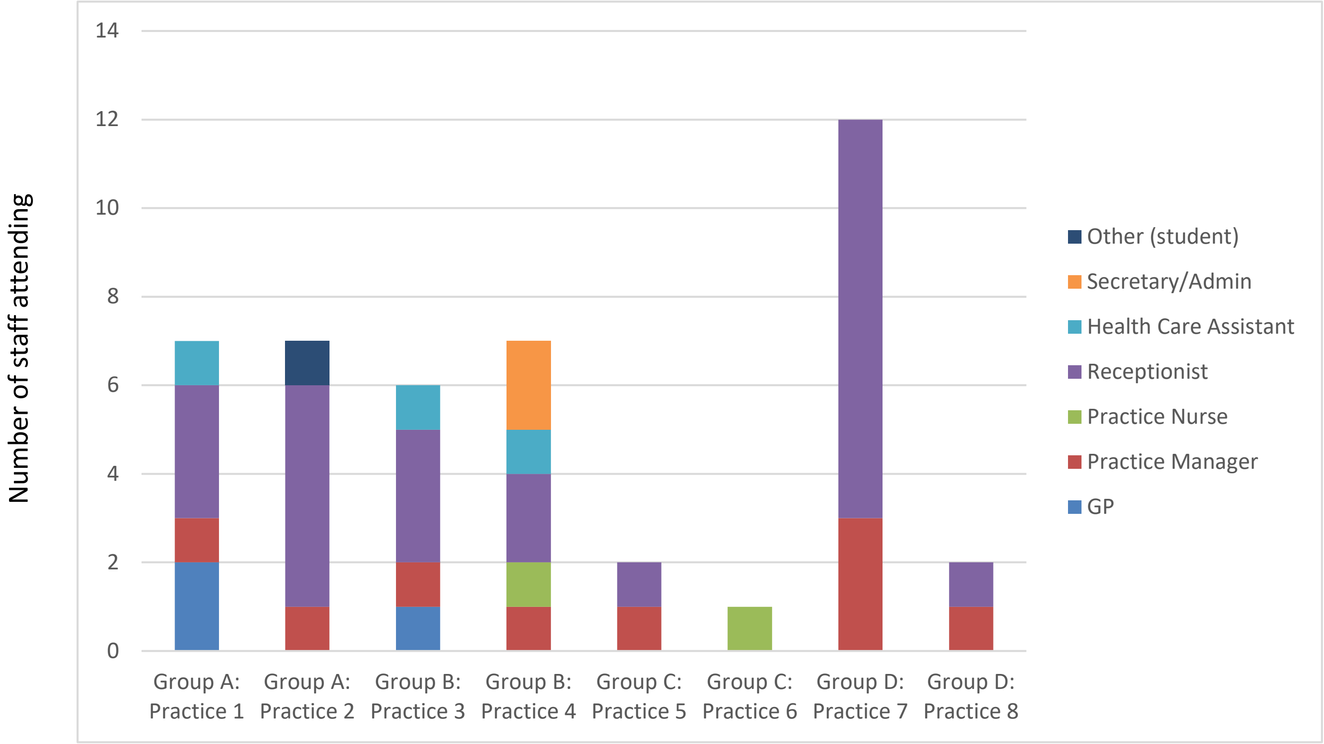


Figure 26. Attendance at RTF set-up sessions by intervention arm, practice and staff role

The total cost to a practice of the system for the 12-week implementation period was £1,144 (including fees paid to the RTF provider). There was no difference in cost between the team-level and individual-level feedback as the processing fee for the report (£75) was assumed to be the same for both.

Facilitated feedback (groups A and B) cost an estimated £678 per practice (SD £227). This comprised £250 (SD £58) in fees to the facilitator and £428 (SD £180) in practice staff time to attend facilitation.

Qualitative evaluation

Staff from four of the participating ten practices (n=8 intervention; n=2 control), took part in focus groups and various staff members from the remaining six practices participated in interviews (n=22). Table 44summarises the characteristics of participating staff.

Table 44. Characteristics of staff who took part in interviews and focus groups

Practice	Intervention group	Number of interviews by staff type	Number of focus group attendees by staff type
1	Facilitated reflection and practice-level feedback		GP(n=2) Administrative including receptionists (n=5) Nurse (n=2)
2	Facilitated reflection and practice-level feedback		GP(n=1) Administrative including receptionists (n=6) Nurse (n=1)
3	Facilitated reflection, practice-level and individual practitioner-level feedback	Deputy practice manager (n=1) Administrative including receptionists (n=2)	
4	Facilitated reflection, practice-level and individual practitioner-level feedback	GP (n=1) Administrative including receptionists (n=2) Nurse (n=1)	
5	Unfacilitated reflection and practice-level feedback		GP(n=1) Administrative including receptionists (n=2) Nurse (n=1)
6	Unfacilitated reflection and practice-level feedback	GP (n=1) Administrative including receptionists (n=1) Nurse (n=1)	
7	Unfacilitated reflection, practice-level and individual practitioner-level feedback	Practice manager (n=1) Deputy practice manager (n=1) Administrative including	

		receptionists (n=2)	
8	Unfacilitated reflection, practice-level and individual practitioner-level feedback	Practice manager (n=1) GP (n=1) Administrative including receptionists (n=2) Nurse (n=1)	
9	Control		GP(n=2) Administrative including receptionists (n=3) Nurse (n=2)
10	Control	Practice manager (n=1) GP (n=2) Administrative including receptionists (n=1) GP registrar (n=1)	

Staff from four practices attended facilitated sessions either during or after the data collection phase, and three facilitators were interviewed once data collection was complete. Researchers visited practices 57 times to conduct observations and patient exit interviews. Out of 375 patients approached for an exit interview, 300 (80%) provided additional comments about their experience of providing real-time feedback. Qualitative data extracted from the RTF devices (patient free text comments) were excluded from the analysis as this information related to patient experience of the practice and not to RTF implementation. Results from the qualitative study are presented under the headings of the four Normalisation Process Theory constructs (coherence, cognitive participation, collective action, reflexive monitoring), sub-divided and illustrated by our findings from the RTF implementation.

Conventions

Sources of quotations are denoted as follows:

- Practice staff
    - 6 digit numeric code denoting individual participant/FG (focus group) or Int (interview)/Staff category
        - PN – Practice Nurse
        - GP – General Practitioner
        - ADM –Administrative staff
        - REC – Receptionist
        - PM – Practice Manager
        - DPM – Deputy Practice Manager
- Patients
    - 6 digit numeric code/PAT
- Facilitators
    - 3 digit numeric code/FAC

Editing is indicated by … where some words are missing, or […] where a larger fragment is missing.

1. Making sense of the RTF implementation (Normalisation Process Theory: ‘Coherence’)

Practice staff used a range of ways to make sense of RTF: by comparing it with other feedback methods; by adapting normal routines to absorb RTF into their practice organisation; by seeing it as part of a programme of communication with their patients; by considering how it may be received by their patient population(s). They drew on their own experience of technological initiatives in other contexts, and of past participation in research studies. Views about the content of the RTF survey influenced overall attitudes towards the concept of receiving immediate feedback.

Feedback and methods

Many members of staff mentioned that they were well-used to receiving feedback from their patients, and that their patients were accustomed to giving feedback. Many made sense of RTF by comparing it to other feedback methods with which they were more familiar, and some favoured the convenience of RTF to traditional paper-based surveys, and mentioned the difficulty of collecting responses:

> *I think it's the way you’ve done it immediately that is much better. Which is if you give them something maybe to take away and bring back or post, they aren't gonna do that … and that’s the end of it ... With the touchscreen it's … much easier.*
>
> *007021/Int/ADM*

Others highlighted problems with obtaining feedback from patients who had not had time to reflect before inputting their responses:

> *You're gonna get some hotheaded responses aren't you? I mean people are going to come out and get really cross, there’ll be some emotion going on there that if they cooled down for five minutes you wouldn’t get those responses.*
>
> *015009/FG/PN*

Some staff voiced concern about access for specific groups of patients. Elderly patients were characterised by some as being less willing or able to use technology. Staff also mentioned patients with low literacy levels and patients whose first language is not English.

Motivation

Some practices were motivated to participate in the study by the requirements of other schemes, such as Friends & Family (CQC) and annual RCGP appraisals.

Context

Many practice staff viewed the RTF implementation purely as part of a research study, and therefore limited in its impact on the practice and workloads:

> *I suppose the key is … in the nicest possible way … having as little impact on sort of patients and staff as possible but gathering enough information for the study to be worthwhile.*
>
> *069027/Int/PM*

For others, RTF was part of an overall strategy for obtaining feedback from patients, and staff valued it as an additional means of staying in touch with their patients' concerns.

Many participants mentioned that people in general are asked for feedback about a range of services, and were able to place RTF within a familiar context:

> *I think one in four people have got smartphones and … I've seen that sort of survey used at airports, it's sort of in … people's lives now really.*
>
> *017019/Int/PM*

RTF content

Staff voiced varying views about the content of the RTF survey. Several were critical about the number of demographic questions. Some were positive about including a couple of their own questions, making the implementation particularly relevant to their practice.

Practice organisation/roles

Practice organisation, including both physical configuration and staff arrangements, affected individual and group perceptions about RTF. Most practices had an electronic check-in system and/or notification screen to call patients to consultations. Sometimes this meant that there was little interaction between receptionists and patients, and reduced opportunities to promote the use of the RTF devices. The position of the RTF devices, often dictated by the physical limitations of the practice building, also influenced levels of use, and the degree to which RTF became embedded within the normal routines of the surgery:

> *It was a bit difficult where it was placed, because we have a booking-in machine and it was next to that, and I think people thought it was another booking-in machine.*
>
> *015019/FG/REC*

Although practices were given materials for involving all staff in the implementation, some viewed it as predominantly a task for the reception team and depended on receptionists to promote the devices to patients:

> *I think we particularly focused on getting reception staff to try and remind patients … as they checked in for their appointment rather than to do it afterwards, cos otherwise it is a lot to ask the clinical staff to remember.*
>
> *011017/FG/PM*

The degree to which practices involved clinical staff varied. In some practices, all staff were familiar with the implementation, so efforts to engage patients were coherent and in tune with the general practice ethos:

> *If they* <patients> *made a comment about the service, I said, 'please can you feed it back'… it was really nice that I could give them something definite to do immediately.*

011001/FG/GP

An individual member of staff, such as the practice manager, or the research champion often took responsibility for managing the implementation. The effectiveness of this individual influenced how RTF was received by both staff and patients. In some practices, where RTF had been explained and promoted successfully, members of staff adapted their existing roles to embrace the new initiative, but in other practices, some individuals voiced discontent and disenfranchisement:

> *As far as I knew it was a kind of if they want to use it, so was it up to me to actually ask them - it wasn't really made that clear.*

069013/Int/REC

Practice knowledge of their patients

Many practice staff used their knowledge of their patients to anticipate how RTF may be received by them. Many felt that particular groups would not be comfortable using a touch-screen device: most often mentioned were elderly patients, patients with literacy problems and patients whose first language is not English.

Staff also mentioned that patients' use of the devices was influenced by processes and volume of work within the practice, both generally and at particular times during the day, week or year:

> *If we're running late, then people were often in a hurry to leave as quickly as they could, having been… held up by us, so that was also an issue I think, for some of the patients.*

010003/FG/GP

In some practices staff concluded that, as their patients were well-used to using touch-screen devices, they would be happy to leave feedback in this way. Others said that patients were not confident or competent with touch-screen technology:

> *I think … giving people access to a computer even if it's a very simple touchscreen computer … is sometimes quite challenging. Just like touchscreen check in, there's a whole group which …hasn't checked in because they've missed one of the buttons.*

021001/Int/GP

2. Working together and with patients to establish RTF (Normalisation Process Theory: 'Cognitive participation and collective action')

The ways in which individuals and groups of staff worked together, and with patients, with regard to RTF varied amongst practices. Findings within these two related NPT constructs (cognitive participation and collective action) are closely linked with some of the sense-making aspects detailed above. Differences in style and methods of communication both within and between staff groups, and with patients, were more pronounced in some practices.

Communication/Differences/Engagement

Staff relationships with each other and with patients were a crucial part of the RTF implementation. Often an open, inclusive approach to communication between staff members coincided with the way in which they related to their patients:

> *There's nothing that's kept away from us … whether it be good or bad … if we have to do something to either make it better or keep up what we're doing then they tell us so being told is the only … way you're gonna provide a service that the patients actually want.*

007021/Int/ADM

By contrast, some administrative staff felt that their knowledge of patients' concerns was ignored and they were not given the opportunity to make suggestions:

> *I mean we can have queues out of the door and it's not noticed by the managers to think, oh right we need more staff there… We are always saying about confidentiality, the patients complain to us a lot about it, but we might mention it and nothing is ever done.*

069013/Int/REC

There were differences in style between individual team members in the same practice:

> *If you want anything filled in you get <name removed> to give it to the patients because she just goes out and says, would you mind filling it in? […] Whereas the others are not quite so interactive with the patients. They'll say something to them over the desk but they won't actually go out and interact with them.*

068015/Int/DPM

Differences between individual patients were often mentioned, and several receptionists admitted that they selected patients to encourage to leave RTF on the basis of their perceived abilities or level of sickness.

Some clinicians were used to asking their patients for feedback, and felt very comfortable doing so, but others were more reticent:

> *I think it's a bit embarrassing though, to say, well could you give me some feedback, I mean I find it quite embarrassing, so I wouldn't ask them.*

010003/FG/GP

One GP suggested that the process of requesting feedback may hint at a hidden agenda within an otherwise positive consultation:

> *It can feel awkward … if the conversation has gone really well, it sometimes slightly undermines the goodness of the conversation or the help that you've given.*

007002/Int/GP

The way in which RTF had been introduced in practices had a profound effect on how engaged individual members of staff felt with the implementation. In practices where communication was inclusive, staff felt part of an important initiative and understood their roles. In other practices, often where the research team had communicated solely with an individual practice contact, some staff felt remote from decision-making and so were not fully engaged with RTF:

> *I haven't been involved. I don't know what the plan is from here.*

016021/Int/REC

Timing and workload

Timing, in many senses, influenced the embedding of RTF in practices, including timing of the request for feedback (after consultation), timing of the implementation itself (in the context of other practice activities), variations in work volume according to particular times and days of the week:

> *It's much easier to get patients to fill things in whilst they're waiting than when they've finished… because they're sitting down waiting, they're almost sitting and looking for something to do.*

011017/FG/PM

Reception staff found fluctuations in workload and demand from patients greatly influenced their ability to play a part in the RTF implementation:

> *Especially on a Monday and a Tuesday when the phones are ringing and the queue is long. It's hard to … explain to them what it's about.*
>
> *068007/Int/REC*

Involving patients

Most practices voiced their intentions to share the results from the RTF devices with their patients via posters/practice website, and several were keen to let their patients know that their responses had led to change

One practice had involved their Patient Participation Group (PPG) in devising additional, practice-specific questions to include in the real-time survey, and another had shared results with the PPG before discussing them as a practice team.

Feedback reports

On the whole practice staff appreciated the regular feedback reports from the study team, and drew favourable comparisons with other surveys, which often entailed in-house data analysis and assembly of results. Many scanned the reports for data which confirmed what they already knew, and some were surprised by the volume of positive feedback.

Several practices were concerned about low RTF response rates, and suggested possible reasons for this, including feedback "fatigue" amongst both patients and practice staff. Some staff believed that patients would not bother to leave feedback unless their experience at the practice was either strongly positive or negative.

3. Appraising and learning from RTF (Normalisation Process Theory: 'reflexive monitoring')

Practice staff viewed the implementation and results in a variety of ways. Some believed that giving their patients an opportunity to leave immediate feedback was a valuable addition to existing ways of communicating with them. Others viewed RTF responses as confirming what they knew already, and possibly could not act on. Many mentioned plans for acting on the RTF, and a few had already instigated changes within the practice by the time they were interviewed or attended focus groups. The degree to which staff trusted the credibility of the results had an inevitable effect on their learning from it, and several staff had suggestions for improving the usefulness of the questions and responses.

Trust/learning/suggestions

Some staff believed that the immediacy of a response added strength to it, but conversely, many felt that a period of reflection was important and could greatly change how the patient viewed their consultation:

> *You might have been denied a medication at the time which might deeply upset you but then two weeks down the line you kind of realise that that's (the) right thing and the feedback might be different.*

007002/Int/GP

Many practice staff noted the low RTF completion rates, and felt that the majority of their patients had not been given the chance to participate:

> *I suppose in the back of your mind you know it's only a small percentage of your whole population, the people who are coming through the door.*

068015/Int/DPM

Some members of staff felt that the RTF devices were used by patients to record two extremes of response, and excluded the middle ground:

> *It attracts two types of people doesn't it, the people who love you and tell you they love you and the people that just had a really bad experience that day and want to take it out on the system, really.*

015016/FG/PM

Many staff found the free text left by patients more useful than the quantitative responses, but some staff found these comments frustrating as they could not follow them up with the individuals concerned:

> *It feels a bit like people might leave feedback if they're unhappy, and so there's very little positive and when you read it back certainly I found it difficult to reflect on and learn from it and to improve my practice because it wasn't very specific.*
>
> *0210017/Int/GP*

Some suggested that an optional facility for respondents to provide their name would be helpful.

Some reception staff mentioned that individual feedback would be helpful as a learning tool for their staff group (the devices were not able to provide individual feedback for receptionists):

> *I take offence at that, cos I think, don't tar us all with the same brush.*
>
> *018011/FG/REC*

Many staff expressed surprise about the positive feedback they received – in both comments and quantifiable data. Negative responses were expected and in many cases confirmed previous feedback from patients.

Action

Several members of staff, predominantly practice managers, mentioned that they had taken or were intending to take action based on RTF. Some were keen to amalgamate the results with data from other initiatives before formulating a plan:

> *What we've historically done is … when we've had a survey, we've published the results of that survey, along with our action plan of how we're going to respond to*

> *different aspects of it and over what timescale … and who would be involved in that … we do do that on a sort of an annual basis […] if you do it in dribs and drabs, it sort of doesn't quite have the same impact.*
>
> *017019/Int/PM*

Some mentioned involving their PPG in discussions about the results and action- planning, whereas others intended to inform their patients by publicising the results in the practice. Some individuals were not sure about what was planned, and did not see it as affecting their own work.

Some staff said that RTF responses had been expected, and covered issues already familiar to the practice, which they had addressed, or were addressing.

Many staff welcomed confirmation of previously-held views, but some felt that their patients' expectations were unrealistic, and it was not possible to satisfy them.

Facilitation

Practice staff had not previously explored patient feedback at sessions guided by an external facilitator. Several factors, occurring before and during the facilitated session, influenced the success of these sessions. These can be summarised as:

(i) *Prior to the session*: communication of the aims and objectives of the session to the entire practice team; dissemination of patient feedback results so that staff could highlight areas for review; and protected time built into staff schedules for reviewing patient feedback.

(ii) *During the session*: skill of the individual facilitator; provision of a clear agenda at the outset including expected outcomes; time to celebrate positive results; and an agreed action plan for staff to work to.

Action planning

Facilitators found it difficult to get staff to agree and commit to a set of actions during the facilitation session. This was partly due to time limitations. Some staff were reluctant to implement changes suggested during the facilitation session, possibly because of previous experience of external pressures:

> *"…a lot of GPs are fairly conservative and don't want to change much, maybe that's because they've got so much change imposed on them, they're reluctant to change"*
>
> *300/FAC*

The facilitation sessions were flexible to allow staff to explore the results, but facilitators suggested that an action plan may have been compiled had it been explicitly included as an intended outcome within the agenda.

Communication and engagement

Some practices were more engaged with the facilitation process than others, and this was demonstrated by their willingness to set a time for staff to meet and discuss RTF. Clear communication of the purpose of the session allowed for staff to contribute to the discussion and share ownership of the decisions being made, especially in practices where facilitation was a novel approach:

> *"But I think that was down to ground setting and me being clear from the start what we are doing. Also explaining the rules of the group and giving them ownership really of what was discussed."*
>
> *002/FAC*

In some practices the reception staff were more engaged with the feedback than clinical staff and provided suggestions for changes to their work routines more readily:

> *"the reception staff seemed fairly willing to contribute …, a lot of the feedback, was very pertinent to them…"*

100/FAC

> *"Yeah, the reception staff are actually better at talking than the GPs because I think they were more enthusiastic (laughter). They were the ones that came up with the ideas."*

300/FAC

Feedback discussion

As mentioned above, in some practices reception and administrative staff made a greater contribution to the discussion than their clinical colleagues. The skills of the individual facilitator had an impact on the discussion; staff from one practice felt 'their' facilitator did not effectively encourage contributions from all staff, but this was not experienced in other practices. Facilitators found it difficult to fully discuss sensitive issues and did not have sufficient time to work through them properly during the session:

> *"You have opened up this box of really quite difficult stuff and then gone away again. What the surgery does with it now is really up to them."*

200/FAC

Time

All facilitators found that the session was not long enough to enable full discussion of feedback, and developing a clear action plan. The timing of the facilitation session was important: the majority took place half-way through the data collection, but some practices commented that it may have been useful at different time intervals during the data collection:

> *"…maybe if we were running it for 6 or 12 months, you could have them at like quarterly intervals…but I think kind of six weeks into it, we were only just getting into it…"*

018015/FG/DPM

Patient Perspectives

Although patients acknowledge the important role of technology, and recognise its value for providing feedback, some highlighted technology as a potential barrier. They mentioned other patients who may be intimidated and deterred from leaving feedback by technology, for example the elderly and those with literacy difficulties. Some patients who chose not to use the RTF device said that they would prefer to feedback directly to the GP or receptionist, and were sceptical about the possibility of changes being made as a result of feedback if not voiced in person to the practice.

The lack of publicity and information about the purpose of the RTF device was a potential barrier for patients and affected their trust that RTF would result in change within the practice. The location of the touch-screen device was an important factor, particularly in terms of privacy. If it was overlooked by reception staff or other patients, individuals often did not use it. Equally, if it was installed in an unobtrusive location, many patients did not notice it.

The timing of the request for feedback was also an important consideration. Some patients said that being asked to stay behind to provide feedback after a consultation was unacceptable, particularly if they had waited a long time for their appointment:

> *"If you are ill you just want to go home after."*
>
> *010001/PAT*

Discussion

Real-time feedback (RTF) is a relatively novel approach to the collection of patient feedback in general practice. In this research, the idea of RTF appeared broadly acceptable to both patients and staff in participating practices. However, communication within the practice team, and between staff and patients, was a key factor that influenced the level of acceptability, and the feasibility, of embedding RTF in practice routine.

Effective communication underpinned the successful implementation of RTF, not only in encouraging patients to use touch-screens in the waiting area, but in the organisation's use of collected feedback. Communication within the practice team influenced staff perceptions of RTF as a useful learning tool and the success of a facilitated reflection session as a means of discussing and planning service change.

In the context of this pragmatic, essentially un-facilitated survey, practice staff and patients viewed RTF positively, but engagement with the touch-screens was lower than reported in other studies from the US. [234, 240] In absolute terms, the majority of practices in the current study collected feedback from 100 or more patients. However, the proportion of consulting patients who used the touch-screens varied across practices (range 0.7% to 8.0%) and, overall, feedback represented the views of a relatively small proportion (mean 3.2%) of consulting patients.

The *absolute* number of patient responses via RTF is comparable to that achieved by the same practices in the most recently published national GP Patient Survey [188] but overall practices' response rates in the national GPPS were much higher (ranging 27% to 53%).

The difference in response rates between the current study and the US studies may reflect the greater number of items in our survey. It may also reflect the lower level of direct encouragement and support provided by staff to help patients use the touch-screens. At many practices, receptionists were given responsibility for encouraging patients to use touch-screens rather than clinicians. Receptionists were observed to interact with a significant proportion of patients who attended the surgery but they were rarely observed to encourage the use of the touch-screens. Whilst a number of reasons were given by patients in the exit surveys for not using the touch-screens, over half of those patients had been unaware of the opportunity to leave feedback; others may have provided feedback if clearer information had been provided about the purpose of the touch-screens. Where staff encouragement to use touch-screens did occur, patients were more likely to start the survey. Direct encouragement was more effective than publicity materials displayed in the waiting area, which went largely unnoticed by patients.

Practices accustomed to collecting and using patient feedback viewed RTF as part of their ongoing dialogue with patients and the immediacy of feedback helped offset the risk of “feedback fatigue” for both staff and patients. However, practices and patients were concerned about patient groups who might be excluded from feedback processes that involve the use of touch-screens, specifically older patients and those for whom English is not a first language. Others felt the RTF screen was easy to read, and acknowledged that people of all ages are well-used to using touch-screen devices in other areas of life. Our analysis suggested that some age groups (46-65 years) were over-represented amongst RTF users, while others were under-represented (over 65 years).

In our study, female patients were more likely to provide RTF than males (62% versus 38% respectively) in contrast to the most recently published national GPPS data, where approximately even proportions of males (49%) and females (51%) responded. [126] The observation of lower rates of feedback in older age groups is in line with that of Dirocco and Day, [234] where more intensive staff support with RTF had been available. Dirocco and Day [234] also reported lower feedback rates among minority ethnic groups. Our study was unable to investigate this as appointments data could not be broken down by patient ethnicity at any of the participating practices.

Our findings with regard to levels of staff engagement with RTF and effective communication within practices and with patients are broadly in line with the earlier UK six-month pilot study. [44] Practice’s physical configuration and flexible assignment of roles can either help or hinder participation and collective action among staff (and with patients) with regard to a new system or process. Good communication about RTF fosters involvement and buy-in from both clinical and administrative staff, including shared reasons for participation, the roles of different staff groups, ongoing progress with RTF collection, and the content of feedback reports. Our findings suggest information was not always communicated effectively to individuals and some felt remote from the process. Wofford et al. [240] suggested that RTF (collected via tablets) had minimal impact on working routines when implemented in a primary care setting. Our findings are more mixed about this: some practices and individuals suspended their involvement with RTF implementation during busy times or with particular patients, while others (particularly clinicians) reported RTF did not impinge on their daily routines.

Practice staff identified potential benefits of using a facilitated session for discussion of patient feedback and having protected time for the celebration of achievements. Effective communication about patient feedback with all members of staff prior to and during a facilitated session encourages constructive debate and all-practice engagement with any changes agreed at the session. Some practices saw advantages in the immediacy of feedback and potential for quick action, in line with existing guidance from the Department of Health [85] and the NHS Practice Management Network. [44] However, other practices preferred to combine their RTF results with other information before considering action, or even action-planning. Many patients commented on the importance of their practice taking account of and acting on feedback, but the degree to which any plans or changes resulting from RTF were communicated to or shared with patients varied greatly.

The costs of RTF need to be compared with outcomes in order to judge whether RTF represents a good investment for a GP practice. Outside the context of a research project, the costs of hiring touch-screens may be borne directly by the practice alongside staff time invested in set-up briefings and team meetings to reflect on patient feedback. GPs and nurses tended only to attend set-up briefing sessions in practices allocated to facilitated feedback, suggesting clinician engagement was higher in those practices. This may be worthy of more detailed investigation in future studies, as it might be a mediator of any observed outcomes. To maximise patient use of touch-screens, consistent effort and time from practice staff (particularly receptionists) is required to directly encourage and support feedback from patients. However, this could be seen as time well-spent if it leads to collection of RTF from a sizeable and representative group of the patient population.

The ability to achieve change in practice is a major issue highlighted in this study. Participants in the qualitative research identified an inertia – perhaps even unwillingness or resistance to implement change - following patient feedback. Such an observation concurs with findings from Deming (2000) who reported on such resistance and inertia: [241]

> *Foreman (Recorded): I fill out a report when anything goes wrong. Someone from management, I was told, would come and take a look at the problem. No one has ever come.*
> *Production worker (Recorded): What good comes of making a suggestion to your foreman? He just smiles and walks away.*
> *And the telling Comment: What else could he do? He does not understand the problem, and could get nothing done if he did*

Strengths and limitations

Our investigation of the acceptability and feasibility of RTF was enhanced by a multi-method approach. A better understanding of the obstacles and drivers associated with embedding RTF in general practices was achieved by organising data from interviews, focus groups and observations according to NPT constructs. Although it is important to note that all four NPT constructs operated and were experienced concurrently, the NPT framework enabled a coherent view of the processes involved in RTF implementation, including the ways in which practice staff and patients understood RTF, teamwork and collective action within practices, and reflection, learning and actions arising from the feedback. Focus groups were attended by a range of staff, and individuals were encouraged to share their views about RTF. Where focus groups were not possible, a range of staff participated in one-to-one interviews.

A range of general practices were recruited to the study, including those in urban, inner city and rural settings, with varying deprivation scores and list sizes. However, practices were drawn from two broad geographical areas (South West and Cambridgeshire) which may not be representative of the UK as a whole. Participating practices may also have been those with an interest in research or service improvement.

The implementation of RTF in this strand of work had inherent limitations when compared to other means of collecting feedback. For example, the survey items were presented only in English, and patients who did not visit the surgery during the implementation period were unable to provide feedback. In some practices, it proved difficult to extract demographic information about consulting patients from the practice system and there was some evidence that appointments data was not consistently recorded within systems, limiting reliable assessment of the response rate and the representativeness of patients who used touch-screens compared to the consulting population. It was not possible to calculate response rates for patients who attended the surgery for reasons other than a consultation. The work undertaken was preliminary in nature, and not intended to address issues relating to overall effectiveness of the RTF intervention, or the related issues pertaining to the timeliness or mode of feedback to practices. Such research would require both considerable additional time and resource to allow for definitive studies to be undertaken.

The implementation in each practice lasted for one 12-week period. In some cases staff noted that they had felt better able to engage with the process because they knew it was time-limited, while others believed that RTF needed more time to become part of the normal routine of the practice. Future studies would need to consider the optimum time period for collecting RTF in general practice, perhaps favouring a more intensive effort to collect feedback for a shorter period of time with the process being repeated after a suitable interval to assess the impact of any resulting service change on patient experience.

Although a key, responsive contact within the practice is an important factor influencing the success of a time-limited research study, spreading information and motivation throughout the practice is crucial. This requires good communication between staff groups and individuals, to foster a sense of involvement at all stages of the implementation, and thereby achieve “buy-in” from the whole practice.

RTF content also needs to be relevant to the concerns of the practice and patients. Some staff were critical of the volume of demographic details required from respondents. Whilst such information was necessary to address the research objectives, it did not reflect the interests of all practice teams. Greater practice and patient involvement with the design and content of the RTF survey may achieve a greater sense of ownership and involvement.

Many of the challenges involved in successfully implementing RTF within practices revolve around the issue of timing. These include issues such as: avoiding “feedback fatigue” (in staff and patients) and duplication of effort by blending RTF with other feedback initiatives; and ensuring teams make contingency plans that take account of busy times within the practice. Consideration also needs to be given to making the best use of patient’s time – for example, patients may have more time and be more willing to use touch-screens to provide feedback about practice services while they are waiting to see a health professional than they are after their consultation.

Conclusions

Despite the low RTF response rate observed when touch-screens were located in general practice waiting areas, patients and practice staff were broadly positive about the concept of real-time feedback. Enhanced buy-in from practice staff and patients might be achieved in a number of ways. This includes involving practices in the design and content of RTF surveys and addressing language barriers and patient concerns about the use of technology. A shared responsibility within the practice to promote and support RTF may result in more proactive encouragement and support of patients to use touch-screen equipment in the waiting area. A longer overall implementation period may be required, during which shorter ‘bursts’ of RTF collection and reporting occur, thus allowing a more thorough assessment of the degree to which RTF can become embedded into general practice and used to improve patient experience. Our reflections on how this work might inform a future trial are outlined in Box 10.

Key learning for planning a future trial of real-time feedback in primary care

- Recruiting and randomising practices to take part in such a study is feasible
- Engaging the whole practice team is of vital importance for the successful implementation of RTF in practice; in particular this requires ensuring engagement and 'buy-in' from staff involved in supporting the day to day delivery of RTF, most commonly reception staff
- It is possible, over time, to attain an acceptable sample size of participating patients, even where full staff 'buy in' has not been achieved
- Thre is a need for focusing effort on securing particpaition from younger and older patients, patient sfrom ethnic minorities, and those with English language difficulties
- RTF based interventions may be costly to implement in practice – at approximately £5 per participant recruited
- Attention needs to be paid to the physical configuration and context of the RTF process
- There is a need for flexible assignment of roles to support RTF implementation and to secure patient participation
- Facilitated feedback is desirable to support RTF in practice, and is welcomed by practice staff
- Multi-methods approaches to evaluation are advantageous

Box 10. Planning a future trial of real-time feedback in primary care

Chapter 11. The validity and use of patient experience survey data in out-of-hours care

This material is reproduced from Warren, F.C., Abel, G., Lyratzopoulos, G., Elliott, M.N., Richards, S., Barry, H.E, et al. Characteristics of service users and provider organisations associated with experience of out of hours general practitioner care in England: population based cross sectional postal questionnaire survey. BMJ 2015;350:h2040 [242] *under an Open Access CC-BY-NC licence. Articles published under CC-BY-NC permit non-commercial use, distribution and reproduction in any medium, provided the original work is properly cited.*

Abstract

Background

In England, out-of-hours general practitioner (GP) services provide urgent medical care to patients when their GP surgeries are closed. National Quality Requirement 5 (NQR5) requires out-of-hours services to routinely audit patient experiences, but provides no guidance on methods. In the absence of comparable data from providers, the out-of-hours items from the national GP Patient Survey have been used to monitor patient experience.

Aims

(i) To explore whether variation in service users' experiences of care were driven by user or provider characteristics; (ii) to document the validity of out-of-hours GP Patient Survey items; and (iii) to understand how providers collect/use patient feedback to drive service improvements.

Methods

A multi-method study, analysing out-of-hours items from the GP Patient Survey dataset (2012/13: 971,232 service users) and a bespoke survey (6 providers; 1,396 services users); and a qualitative interview study with staff (11 providers, 31 staff).

Findings

Service users provided less positive ratings of out-of-hours care provided by commercial organisations when compared with those using either NHS or not-for-profit providers; service users whose ethnic origin was 'non-white', or those finding it difficult to take time off work to attend their general practice also reported poorer experience. GP Patient Survey data,

subject to minor modifications, appeared valid and thus suitable for benchmarking. However, the items need updating to reflect the changes made to accessing out-of-hours services by telephone. Patient feedback (including GP Patient Survey) has a limited role in driving changes to out-of-hours service provision, in part due to the lack of clarity of NQR5.

Conclusions

Out-of-hours items on the GP Patient Survey require refinement, but appear suitable for benchmarking purposes. NQR5 is ambiguous and requires revision to assist providers in collecting and acting upon patient feedback.

Introduction and rationale

Defining out-of-hours GP care

In England, out-of-hours GP services provide urgent medical care to patients when their GP surgeries are closed i.e. between 6.30pm and 8.00am on weekdays, and at weekends and bank holidays. Although medical care is largely provided by GPs, nurses and emergency care practitioners may also provide clinical care. Out-of-hours services are provided to manage health problems that cannot wait until the next working day; these services are not intended as an alternative route to healthcare for non-urgent problems for those patients who cannot attend during practice opening hours. Recent English national audit data reported that out-of-hours GP services handled around 5.8 million contacts during 2013-14, of which 3.3 million were face-to-face patient consultations. [243]

The provision of out-of-hours GP care has changed significantly over the last decade in England. In 2004, responsibility for out-of-hours services transferred from local GPs to NHS primary care commissioners. Commissioners are now responsible for purchasing care from provider organisations and, in some regions, very different models of care have emerged. In England there are currently 211 Clinical Commissioning Groups [243] commissioning out-of-hours services, although the number of providers is smaller, as many providers contract with two or more neighbouring commissioners. Out-of-hours services are also provided by different types of organisations, including NHS Trusts, not-for-profit providers (for example, social enterprises), and commercial healthcare providers. [62, 244] Such services continue to evolve; since the phased introduction of the NHS 111 service, completed in February 2014, [243] different providers may provide different aspects of care, for example call-handling and delivery of clinical care, in the same geographical area.

Ensuring quality and safety of out-of-hours care

Although the re-organisation of out-of-hours GP care has the potential to bring about new approaches and increased efficiency of service provision, such reconfiguration may also generate reduced service coverage or quality. To tackle concerns regarding the quality of care provided, national standards were published with which all out-of-hours GP care providers were expected to comply. [63] Providers are required to report their performance to their commissioners across a range of National Quality Requirement (NQR) recommendations. Of particular relevance to the IMPROVE programme is recommendation 5 (NQR5), which mandates out-of-hours providers to regularly audit a random sample of patients' experiences and to take appropriate action on the results.

Despite the introduction of the NQRs, criticism of the quality and safety of out-of-hours care persists. [62, 245] Prompted by the death of a patient in 2008, the Care Quality Commission (CQC) investigated the case and produced additional recommendations for commissioners and providers of out-of-hours services regarding performance assessment. [244] More widely, urgent care provision in England has been criticised regarding service accessibility, the lack of continuity of care and concerns about patient safety.[246-249] Within this context, the CQC has recently assumed responsibility for regulating and inspecting the quality and safety of out-of-hours primary care services. [250] With CQC inspections commencing October 2014, the latest CQC overview reported the majority of service provision to be of high quality, but with some areas where improvements could be made. [251]

Role of patient experience surveys in quality assessment

From 2015, service commissioners will be expected to publish annual data on provider performance against the NQRs. [252] Such a requirement is problematic for NQR5, as there is no agreed methodology for conducting patient experience audits. Without reliable and valid methods of assessing patient experience, it is impossible for providers to accurately assess their own performance and to subsequently use this information to guide service improvement. Providers may also use different tools and survey methods, and the resultant data cannot be used for the purposes of benchmarking to assess variations in service quality between providers. Although a number of standardised patient questionnaires are available to assess patient experiences of out-of-hours primary care services, [253] these tools have not been widely adopted in routine practice.

While it is not possible to benchmark out-of-hours providers using the patient experience data collected for NQR5, the 2014 national audit of GP out-of-hours care [243] and CQC both analysed patient experience data from the English GPPS. The GPPS includes six items relating to out-of-hours care (two 'access' and four 'evaluative' items). As the only large-scale population survey of patients' understanding, use and experiences of out-of-hours care benchmarking of GPPS data is potentially possible. Establishing the validity of the GPPS out-of-hours items is, however, an important prerequisite to using this data to document variation in scores between out-of-hours services and for benchmarking. We have previously published evidence to support the reliability of the GPPS (including out-of-hours items). [133] Using a range of different methods and analytic approaches, we have also demonstrated the validity of GPPS items evaluating in-hours primary care services, [120, 254] but this has yet to be established for out-of-hours care items.

Once the causes of poor patient experience of out-of-hours care have been understood, interventions to improve care can then be designed. However, the current literature on the effects of feedback of patient assessments is insufficient in scope, quality and consistency to design effective interventions targeting service delivery and organisation, or the performance of clinicians. [18, 76, 255]

Rationale for the out-of-hours research

This research was designed to address these gaps in our knowledge to enable managers, patients and professionals to have confidence in the meaning of patient assessments of out-of-hours primary care services recorded in the national GPPS. The work package addressed three important areas:

The **first workstream** built on earlier analysis of the GPPS, which reported that important socio-demographic variations exist in patient experiences of in-hours primary care services, [120] but did not examine if such variations existed for out-of-hours items. Given that the CQC and National Audit Office have both used GPPS to monitor service users' experience of out-of-hours care, it is important to understand whether variation in service users' experiences of care is driven by user characteristics, as opposed to differences in the care provided by different types of providers.

The **second workstream** sought to explore the validity of the out-of-hours items from the GPPS. The Out-of-hours Patient Questionnaire (OPQ) is a complementary tool to the GPPS, which collects more detailed information on patient experience of out-of-hours care, and has undergone more extensive testing and validation. [8, 83, 256] The second project tested the performance of GPPS out-of-hours questions against data derived from the OPQ to examine the validity of GPPS items.

The **third workstream** examined how out-of-hours GP services make sense of the information provided by patient questionnaires and, where possible, use this information to design interventions to improve patient experience through service reconfiguration and development.

Structure of the out-of-hours work package

The out-of-hours work package consisted of three workstreams, each of which used different datasets and methods. The remainder of this chapter describes the study aims and objectives, methods, results and discussion arising from each of the three workstreams in turn, before summarising the key conclusions that arose from the work programme.

Stakeholder advisory group

A stakeholder advisory group composed of three representatives from out-of-hours service providers, two primary care academics and a service user was convened to support Workstreams 2 and 3. The group met to review study methods and procedures in light of the findings of preliminary piloting and testing of methods (see Workstream 2), and to comment on topic guides supporting interviewing in Workstream 3. Due to the logistical challenges of organising face-to-face meetings around staff availability, after an initial face-to-face meeting most advisory group input was secured by email communication and telephone.

The original aim was to recruit two service users through our links with local service providers, and using methods recommended by our Exeter University supported Public and Patient involvement groups (http://clahrc-peninsula.nihr.ac.uk/patient-and-public-involvement-in-research and http://www.folkus.org.uk). Potential service user participants were provided with a brief information sheet regarding what would be involved in advisory group membership, and were informed that any costs incurred preparing or attending advisory group meetings would be reimbursed. Despite significant efforts to secure lay stakeholder participation, it proved difficult to recruit service users with relevant, lived experience to the advisory board. Whilst difficulties in recruiting lay advisors was problematic to the research, provider staff members indicated that their services experienced similar problems, probably due to the nature by which patients consulted (i.e., relatively infrequent consulters seeking care for an urgent problem) and the lack of continuity between provider and service user.

Changes to study methods from the original protocol

The overall aim of this strand of work, as stated in the original protocol, was:

“To investigate how the results of the GP Patient Survey can be used to improve patients’ experience of out-of-hours care (aim 7).”

In our original application, we specified four objectives within this, three of which were successfully addressed within this programme (objective 1: cognitive testing of GPPS out-of-hours items; objective 2: establishing GPPS item validity and reliability; and objective 3: identifying how data from GPPS can be effectively used to inform out-of-hours service reconfiguration). Objective 4, undertaking preliminary piloting of an intervention to improve patient experiences of out-of-hours care, was not achieved. The qualitative research undertaken to address aim 3 identified significant heterogeneity in terms of how providers collected and acted upon patient feedback, and of the perceived utility of GPPS as a platform on which to mount quality improvement. It was clear upon completion of qualitative work with service providers that more research was needed to design, and then test the feasibility and acceptability of an intervention to embed patient feedback within quality improvement cycles.

For the three objectives that were achieved, some minor modifications to study methods were implemented as the full protocols were developed. For example, it was initially proposed to interview up to 45 patients to test user responses to out-of-hours GPPS items. In reality, only 20 service users underwent cognitive interviewing, as this proved sufficient for testing the validity of items. Similarly, to address objective 3, a more ambitious, qualitative interview study was undertaken with staff from out-of-hours services. Here eleven English providers (rather than six) were sampled and interviewed to ensure greater diversity in the types of provider organisation, and the populations served.

Workstream 1. Exploring variations in national GP Patient Survey out-of-hours items

Study aims and objectives

This workstream investigated:

- potential associations between service users’ evaluations of out-of-hours GP care and individual level socio-demographic factors;
- whether variations in evaluations were related to `clustering’ of service users reporting poorer experience within providers reporting poorer performance overall; and
- whether there was an association between service users’ evaluations and type of provider organisation (NHS, commercial or not-for-profit organisations).

To address these aims, an analysis of service users’ ratings of out-of-hours GP care from GPPS data was undertaken.

Methods

Patient questionnaires

GPPS data (July-September 2012 and January-March 2013) were analysed (overall response rate of 35% (971,232/2,750,000). [183] The GPPS included four evaluative questions on out-of-hours provision, three of which were analysed: ‘timeliness’ of receiving care (‘about right’, ‘took too long’, or ‘don’t know/doesn’t apply’), ‘confidence and trust’ in the out-of-hours clinician (‘yes, definitely’; ‘yes, to some extent’; ‘no, not at all’, or ‘don’t know/can’t say’), and ‘overall experience’ of the out-of-hours GP service (5-point Likert scale from ‘very good’ to ‘very poor’). These questions were only completed by service users who had attempted to contact an out-of-hours GP service within the preceding six months.

Service user characteristics

Five socio-demographic variables derived from GPPS responses were analysed: gender (male as reference); ethnicity (White (reference), versus five categories derived from Office for

National Statistics [257]); age in eight categories (18–24 years as reference); parent status (non-parent (reference)); and whether the service user was able to take time away from work to attend their practice during working hours (individuals ‘not in paid work’ (reference), ‘paid work, can take time away’ or ‘paid work, could not take time away’). A sixth socio-demographic variable, deprivation (national Index of Multiple Deprivation (IMD) fifths; ‘least deprived’ as reference) was determined based on the respondent’s residential postcode. [258]

Practice and out-of-hours GP service providers

Each service user was mapped to the out-of-hours GP provider responsible for providing clinical care for the service user’s practice during the six month period prior to sending the questionnaire. Mapping was achieved for 96% (934,931/971,232) of service users in the dataset; 7886 practices were mapped to 91 out-of-hours GP providers, of which 86 had an identifiable provider organisation type (not-for-profit (reference), NHS or commercial).

Statistical methods

Analyses were performed using Stata v.12 (StataCorp, Texas, USA). Socio-demographic data are described for all service users contacting an out-of-hours GP provider in the previous 6 months (for themselves or on behalf of another person). To facilitate comparison between measures on different scales, outcomes were linearly rescaled from 0–100, [209] with a difference of <3 points considered ‘small’ in respect of practical significance. [94] Missing data at the level of service users or providers (including ‘don’t know’/’does not apply’) were excluded from analysis. It was assumed that service user responses would be `clustered’ by out-of-hours provider (not practice) with clustering adjusted as a random effect.

Three statistical models were employed. Model A comprised a fixed effect multivariable linear regression model including individual socio-demographic factors as covariates, and generated mean differences in outcome scores for comparator socio-demographic groups compared with reference categories, without accounting for differences in outcome across providers. Model B was a mixed effects model which extended Model A by incorporating a random intercept for provider. Model B therefore adjusted for differences in outcome between providers, and estimated the mean difference between comparator and reference group in outcome scores within providers. Comparing Models A and B identified the extent to which any overall difference between service users of specific socio-demographic groups was due to clustering of service users within providers achieving a low outcome score. [209]

Model C extended Model B by adding ‘provider type’ as a covariate. This model estimated the effect of provider type, with adjustment for service user characteristics, for each outcome. Comparing the between-provider variance from Models B and C quantified the degree of between-provider variation attributable to provider type. The effect of provider type, analogous to an effect size such as Cohen’s *d*, was expressed as the standardised mean difference (mean difference between comparator provider type and not-for-profit providers divided by the between-provider SD derived from Model C).

Results

The socio-demographic characteristics of 106,513 service users (from 7492 practices) who had contacted an out-of-hours provider, and were mapped to a provider of a known organisation type, are shown in Table 45. Service users’ overall evaluations of out-of-hours GP services were generally positive (Table 46); 71% (73,983/103,523) of participants reported a ‘very good’ or ‘fairly good’ overall experience, although 31% (31,966/104,145) felt it took too long to receive care.

Table 45. Socio-demographic characteristics of service users contacting an out-of-hours GP provider (on their own behalf or for someone else)

Socio-demographic characteristics[1]	**N=106,513**
Gender; n (%)	
Male	38,553 (36.6)
Female	66,879 (63.4)
Total	105,432
Age; n (%)	
18–24	4850 (4.6)
25–34	14,745 (14.0)
35–44	20,066 (19.0)
45–54	18,699 (17.7)
55–64	16,760 (15.9)
65–74	14,704 (13.9)
75–84	11,201 (10.6)
85 and over	4509 (4.3)
Total	105,534
Ethnic group; n (%)	
White	90,034 (85.5)
Mixed/multiple ethnic groups	860 (0.8)
Asian/Asian British	7985 (7.6)
Black/African/Caribbean/Black British	2471 (2.4)
Other ethnic group	3934 (3.7)
Total	105,284
Deprivation fifth; n (%)	
1 (least deprived)	19,537 (18.4)
2	20,672 (19.4)
3	21,633 (20.3)
4	21,486 (20.2)
5 (most deprived)	23,028 (21.7)
Total	106,356
Parent/guardian of children aged under 16? n (%)	
No	61,276 (62.8)
Yes	36,277 (37.2)
Total	97,553
Can you take time away from work to see a GP during your typical working hours? N (% of total; % of total relevant)	
Not relevant[2]	51,027 (51.3; NA)
Yes	31,298 (31.5; 64.7)
No	17,057 (17.2; 35.3)
Total	99,382
Total relevant	48,355

[1] Participants mapped to 86 out-of-hours providers via 7492 practices.

[2] Service users reported that they were not doing paid work (e.g. retired, unemployed, full-time student). NA: not applicable.

Reproduced from Warren, F.C., Abel, G., Lyratzopoulos, G., Elliott, M.N., Richards, S., Barry, H.E, et al. 2015 under an Open Access CC-BY-NC licence. [242]

Table 46. Timeliness of care, confidence and trust in out-of-hours clinician, and overall experience of care: raw scores

Question		**Response frequency[1] n (%)**
How do you feel about how quickly you received care from the out-of-hours GP service?		
	It was about right	65 298 (62.7)
	It took too long	31 966 (30.7)
	Don’t know/doesn’t apply	6881 (6.6)
	Total	104,145
Did you have confidence and trust in the out-of-hours clinician you saw or spoke to?		
	Yes, definitely	42,264 (40.7)
	Yes, to some extent	42,938 (41.3)
	No, not at all	12,222 (11.8)
	Don’t know/can’t say	6490 (6.3)
	Total	103,914
Overall, how would you describe your experience of out-of-hours GP services?		
	Very good	33,662 (32.5)
	Fairly good	40,321 (39.0)
	Neither good nor poor	15,638 (15.1)
	Fairly poor	8140 (7.9)
	Very poor	5762 (5.6)
	Total	103,523

[1] Includes all service users mapped to an out-of-hours GP provider with known organisation type even if complete demographic data not available.

Reproduced from Warren, F.C., Abel, G., Lyratzopoulos, G., Elliott, M.N., Richards, S., Barry, H.E, et al. 2015 under an Open Access CC-BY-NC licence. [242]

Data were included for 86 providers: 44 not-for-profit, 21 NHS, and 21 commercial. Provider type was associated with all three outcomes (global *p*-value <0.001 for ‘trust and confidence’ and ‘overall experience’, *p-value* = 0.013 for ‘timeliness’). No statistically significant differences were observed between NHS and not-for-profit organisations with regard to any of the outcomes, whilst commercial providers scored lower than not-for-profit organisations for all three outcomes (Table 47). The magnitude of these differences was approximately 3 points (Model C) for all outcomes.

Table 47. Associations of out-of-hours GP provider type with 'timeliness', 'confidence and trust', and 'overall experience' of care

Provider type[1]	**Mean difference**[2] **(95% CI)**	***p*-value**[3]	**Standardised mean difference**
Timeliness of out-of-hours GP care			
	Model C[4,5] (N providers=86; N service users=83,176) Between provider standard deviation 5.19		
NHS	1.28 (-1.61 to 4.17)	0.013	0.25
Commercial	-3.52 (-6.40 to -0.64)		-0.68
Confidence and trust in out-of-hours clinician			
	Model C[4,5] (N providers=86; N service users=83,316) Between provider standard deviation 3.14		
NHS	1.00 (-0.79 to 2.79)	<0.001	0.32
Commercial	-3.25 (-5.03 to -1.46)		-1.04
Overall experience of out-of-hours GP care			
	Model C[4,5] (N providers=86; N service users=88,423) Between provider standard deviation 3.33		
NHS	1.07 (-0.77 to 2.90)	<0.001	0.32
Commercial	-3.13 (-4.96 to -1.30)		-0.94

[1] Models included 44 not-for-profit providers (reference group), 21 NHS providers, and 21 commercial providers.
[2] All outcomes linearly rescaled from 0 to 100.
[3] *p*-value refers to global effect of covariate across all groups vs. reference group.
[4] Models adjusted for age, gender, ethnic group, deprivation, parent status and ability to take time away from work during work hours.
[5] Random effect on provider of out-of-hours GP provider organisation.
Reproduced from Warren, F.C., Abel, G., Lyratzopoulos, G., Elliott, M.N., Richards, S., Barry, H.E, et al. 2015 under an Open Access CC-BY-NC licence .[242]

A comparison of the between-provider variance (Model B versus C) for 'overall experience of care' observed that 18.6% of the between-provider variability was due to provider type (Table 47, Table 48), 11.3% for 'timeliness' (Table 47, Table 49) and 20.9% for 'confidence and trust' (Table 47 and Table 50). The standardised mean difference for commercial provider type compared with not-for profit was -0.68 SDs for 'timeliness', -1.04 SDs for 'confidence and trust', and -0.94 SDs for 'overall experience' outcomes (Table 48). This equates to a moderate ('timeliness') or large ('confidence and trust', 'overall experience') effect size attributable to commercial provider type.

Table 48. Overall experience of out-of-hours GP services: linear regression modelling

Socio-demographic covariate	**Model A[1,] (N=88,423) Overall difference**		**Model B[1,2] (N providers=86; N service users=88,423) Within out-of-hours provider difference Between-provider standard deviation 3.69**		**Percentage of overall difference (if negative) attributable to clustering of socio-demographic group in lower scoring providers**
	Mean difference[3] (95% CI)	*p*-value[4]	Mean difference (95% CI)	*p*-value[4]	
Ethnic group[5]					
Mixed	-3.44 (-5.47 to -1.41)	<0.001	-2.01 (-4.03; 0.01)	<0.001	42
Asian	-5.61 (-6.32 to -4.90)		-3.62 (-4.36; -2.89)		35
Black	-2.14 (-3.40 to -0.89)		0.13 (-1.14; 1.40)		>100
Other	-0.75 (-1.78 to 0.27)		1.29 (0.25; 2.32)		>100
Able to take time away from work during typical working hours[6]					
Yes	1.30 (0.82 to 1.78)	<0.001	1.29 (0.81 to 1.76)	<0.001	Not applicable
No	-4.79 (-5.36 to -4.23)		-4.73 (-5.29 to -4.17)		1

[1] Models also adjusted for age, gender, deprivation, and parent status. [2] Random effect on out-of-hours GP provider organisation. [3] All outcomes linearly rescaled from 0 to 100. [4] *p*-value refers to global effect of covariate across all categories versus the reference category. [5] Mixed: mixed/multiple ethnic groups. Asian: Asian/Asian British; Black: Black/African/Caribbean/Black British; Other; reference: White. [6] Reference group (Not relevant) includes service users who responded to the question “Which of these best describes what you are doing at present?”: full-time education; unemployed; permanently sick or disabled; fully retired from work; looking after the home; doing something else

Reproduced from Warren, F.C., Abel, G., Lyratzopoulos, G., Elliott, M.N., Richards, S., Barry, H.E, et al. 2015 under an Open Access CC-BY-NC licence .[242]

Table 49. Timeliness of care from out-of-hours GP services: linear regression modelling

Socio-demographic covariate	**Model A[1,] (N=83,176) Overall difference**		**Model B[1,2] (N providers=86; N service users=83,176) Within out-of-hours provider difference Between-provider standard deviation 5.51**		**Percentage of overall difference (if negative) attributable to clustering of socio-demographic group in lower scoring providers**
	Mean difference[3] (95% CI)	***p*-value[4]**	**Mean difference[3] (95% CI)**	***p*-value[4]**	
Ethnic group[5]					
Mixed	-4.78 (-8.34 to -1.23)	<0.001	-3.45 (-6.99 to 0.09)	<0.001	28
Asian	-13.27 (-14.51 to -12.03)		-11.08 (-12.37 to -9.79)		17
Black	-7.64 (-9.86 to -5.42)		-5.67 (-7.92 to -3.42)		26
Other	-8.44 (-10.24 to -6.64)		-6.57 (-8.40 to -4.75)		22
Able to take time away from work during typical working hours[6]					
Yes	3.45 (2.62 to 4.27)	<0.001	3.48 (2.65 to 4.30)	<0.001	Not applicable
No	-6.58 (-7.56 to -5.61)		-6.48 (-7.45 to -5.51)		2

[1]Models also adjusted for age, gender, deprivation, and parent status. [2]Random effect on out-of-hours GP provider organisation. [3]All outcomes linearly rescaled from 0 to 100. [4]*p*-value refers to global effect of covariate across all categories versus the reference category. [5]Mixed: mixed/multiple ethnic groups. Asian: Asian/Asian British; Black: Black/African/Caribbean/Black British; Other; reference: White. [6] Reference group (Not relevant) includes service users who responded to the question "Which of these best describes what you are doing at present?": full-time education; unemployed; permanently sick or disabled; fully retired from work; looking after the home; doing something else.

Reproduced from Warren, F.C., Abel, G., Lyratzopoulos, G., Elliott, M.N., Richards, S., Barry, H.E, et al. 2015 under an Open Access CC-BY-NC licence. [242]

Table 50. Confidence and trust in out-of-hours clinician: linear regression modelling

Socio-demographic covariate	**Model A[1,] (83,316) Overall difference**		**Model B[1,2] (N providers=86; N service users=83,316) Within out-of-hours provider difference Between-provider standard deviation 3.53,**		**Percentage of overall difference (if negative) attributable to clustering of socio-demographic group in lower scoring providers**
	Mean difference[3] (95% CI)	*p*-value[4]	Mean difference[3] (95% CI)	*p*-value[4]	
Ethnic group[5]					
Mixed	-3.02 (-5.58 to -0.46)	<0.001	-1.72 (-4.27 to 0.84)	<0.001	43
Asian	-5.95 (-6.85 to -5.05)		-3.92 (-4.86 to -2.99)		34
Black	-2.62 (-4.22 to -1.02)		-0.33 (-1.95 to 1.29)		88
Other	-1.18 (-2.48 to 0.13)		0.87 (-0.46 to 2.19)		>100
Able to take time away from work during typical working hours[6]					
Yes	2.24 (1.64 to 2.84)	<0.001	2.23 (1.63 to 2.82)	<0.001	Not applicable
No	-5.35 (-6.05 to -4.64)		-5.27 (-5.97 to -4.57)		1

[1] Models also adjusted for age, gender, deprivation, and parent status. [2] Random effect on out-of-hours GP provider organisation. [3] All outcomes linearly rescaled from 0 to 100. [4] *p*-value refers to global effect of covariate across all categories versus the reference category. [5] Mixed: mixed/multiple ethnic groups. Asian: Asian/Asian British; Black: Black/African/Caribbean/Black British; Other; reference: White. [6] Reference group (Not relevant) includes service users who responded to the question "Which of these best describes what you are doing at present?": full-time education; unemployed; permanently sick or disabled; fully retired from work; looking after the home; doing something else.

Reproduced from Warren, F.C., Abel, G., Lyratzopoulos, G., Elliott, M.N., Richards, S., Barry, H.E, et al. 2015 under an Open Access CC-BY-NC licence. [242]

Service users of mixed ethnicity and Asian ethnicity reported poorer care for all three outcomes compared with White respondents; a more variable pattern of care was evident for service users of black ethnicity and other ethnicity (Tables 48-50). In general, the mean differences in scores between White service users and service users from the mixed, black and other ethnic groups tended to be of lower magnitude that those between Asian and White service users.

A comparison of models A and B indicated that with regard to ‘timeliness’, only 17% of the mean difference in scores between Asian and White service users derived from Model A (-13.27, 95% CI -14.51 to -12.03; Table 45) was due to clustering of Asian service users within providers that scored lower overall (versus 28%, 26% and 22% in mixed, black and other ethnicity services users respectively). For overall experience of care, 35% of the mean difference between Asian and White service users (-5.61, 95% CI -6.32 to -4.90; Model A, Table 44) was attributable to clustering of Asian service users within a lower scoring provider.

Service users who could not take time away from work to attend their practice reported lower mean scores across all three outcomes when compared to those where this was not applicable, whereas service users who could take time away from work reported higher mean scores (Tables 48-50)

Other individual level socio-demographic characteristics (gender, age, deprivation and parent status) were also associated with the three outcomes measures (deprivation was only associated with ‘trust and confidence’ and ‘overall experience’), but the effects were not explored further due to the small magnitude of the mean differences when compared with the relevant reference category, or due to more positive scores in the comparator category (i.e. potentially more disadvantaged) than in the reference group.

Discussion

Analysis of GPPS data identified that commercial provider organisations were associated with poorer reports of care across all three outcome measures when compared with not-for-profit organisations after controlling for patient-level socio-demographic characteristics. The lower scores associated with commercial providers is consistent with observations from USA data that found that for-profit hospitals were associated with worse patient experience than non-profit hospitals. [259, 260] However, the reasons underlying the lower scores for commercial organisations, even after controlling for individual socio-demographic variables, are unclear. This may reflect a genuine poorer experience of care provided by commercial providers, or the willingness of commercial providers to operate in areas deemed less attractive to NHS or not-for-profit organisations. It may also be that service users' perceptions of provider type influenced their ratings, although it is questionable whether service users are aware whether their provider was commercial as opposed to NHS or not-for-profit, except perhaps in areas where media attention has focused on their local service.

Service users from minority ethnic groups tended to report less favourable care than White service users, with some variation observed across out-of-hours providers. This finding was in part attributable to clustering of minority ethnic service users in out-of-hours GP services with lower overall scores. Previous analysis of GPPS regarding 'in-hours' care has indicated that minority ethnic patients reported generally lower experience scores, [209] and that patients of different ethnic backgrounds may differ with regard to drivers of satisfaction. [94] In our analyses, although Asian service users reported lower mean scores than White service users for all three experience outcomes, the greatest difference was regarding the timeliness of care. Similar differences were seen for other ethnic groups, but of a lesser magnitude, suggesting that service users from minority ethnic groups, and Asian service users in particular, place substantial value on the timeliness of out-of-hours care. The ability of an out-of-hours GP service to meet service users' expectations has previously been argued to be a strong driver of satisfaction with care [261], although this cross-sectional analysis cannot definitively answer this question.

An inability to attend the practice due to work commitments was also significantly associated with lower scores across all three outcomes compared with those not in paid work, while individuals who reported being able to take time off work reported somewhat better experiences. One explanation is that out-of-hours providers, who do not provide routine `non-urgent' care, may not meet the expectations of service users who find it difficult to attend their practice during regular hours. However, as no information on the nature or the urgency of the service user’s health condition was available this question cannot be addressed definitively.

Strengths and limitations

Unlike CQC and national audit data, this analysis of GPPS data was the first to map the majority of practices (and hence service users) to a specified out-of-hours GP provider, and to determine the organisational provider type. The large sample available enabled sophisticated modelling to test the associations between provider and service user socio-demographic characteristics, and service user evaluations of care.

Several limitations were evident regarding the data available from the GPPS. Service users were invited to provide feedback on their experiences of out-of-hours care in the preceding 6 months. Recall bias cannot be discounted, as previous research has found that that older patients may not accurately report health service resource use over the short time frame of 3 months. [262] No data were collected regarding the nature/urgency of the service user’s complaint, the time/date of the contact, or how the contact was managed. Although data on ethnicity were collected, the GPPS did not ask about service users’ English language ability, nor about educational attainment, both of which may be related to experience of care. [73] The lack of detailed response options regarding whether the service user was able to take time away from work, and timeliness of care also restricted our ability to interpret these data.

The GPPS response rate of 35% is also problematic. However, no evidence of adverse association between response rate and non-response bias has been found for the GPPS, and previous research using rigorous probability sampling methods (as used in the GPPS) have only observed a weak association between non-response rates and non-response bias. [125, 173, 263] An analysis of data on out-of-hours care in the Netherlands suggested that non-response bias was small in respect of overall satisfaction with out-of-hours care. [264]

Workstream 2. Establishing the validity of GP Patient Survey out-of-hours items

Study aims and objectives

The overarching aim was to establish the validity of the GP out-of-hours care items within the GPPS to inform its suitability for benchmarking providers. This was achieved through a multi-method project composed of two stages. In the first stage, preliminary psychometric testing of the out-of-hours items was undertaken through cognitive interviews, combined with a pilot survey of out-of-hours users to test survey methods. The second stage tested the hypothesis that the GPPS items (modified after piloting) would demonstrate construct validity if together the GPPS items were correlated with the two known subscales of the OPQ (an established, valid and reliable measure of patient experience. [8, 256] Concurrent validity would be established if the thematically-relevant OPQ items were found to be associated with each of the GPPS items in linear regression modelling.

Methods

Settings

Six out-of-hours providers across England were recruited for a cross-sectional survey of service users. Data from Year 5, Quarter 2 (July-September 2010) GPPS (www.gp-patient.co.uk) was used to sample providers to ensure there was variation in respect of performance (high/medium/low scoring) on respondents' overall ratings of care received by GP out-of-hours services, as well as the type of provider (NHS, commercial, social enterprise), and the geographical area covered by the service (inner city/suburban, rural). Two participating service providers were operated by NHS Trusts, three were operated by commercial companies, and one was a not-for-profit social enterprise.

Survey piloting and cognitive interviews

A pilot study was conducted with two providers; distributing study questionnaires to 500 service users (*n*=250 per provider). Cognitive interviews with out-of-hours service users were conducted to explore the cognitive challenges faced by service users when completing the GPPS out-of-hours items and establish validity of the item set. Twenty service users (predominately female and aged 65 years or older), from two out-of-hours providers, were interviewed using a think aloud and four-stage verbal probing approach. [265] Interviews were audio-recorded, transcribed verbatim and analysed using protocol analysis. [265]

This preliminary work highlighted issues with GPPS questions and with sampling of service users. The GPPS filters respondents to the out-of-hours items if they report having tried to make contact with a GP out-of-hours service in the previous six months, either for themselves or someone else. Since the respondents for this study were sampled from known users of out-of-hours providers, respondents were requested to evaluate their experience of the last time they made contact with a GP out-of-hours service. Minor modifications to the wording of the GPPS out-of-hours items (1 item) and/or response options (Table 51) and sampling exclusion criteria were suggested by the study team. These changes were reviewed and approved by the study advisory group prior to commencing data collection.

Table 51. Changes made to GP Patient Survey items evaluating out-of-hours care following cognitive interviews with service users

GPPS item wording	GPPS response options	Revised wording	Revised response options
Q38. How easy was it to contact the out-of-hours GP service by telephone?	Very easy; fairly easy; not very easy; not at all easy; don't know/ didn't make contact	No changes made	Very easy; fairly easy; not very easy; not at all easy; don't know / didn't make contact by telephone
Q39. How do you feel about how quickly you received care from the out-of-hours GP service?	It was about right; it was too long; don't know / doesn't apply	No changes made	It was quicker than expected; it was about right; it was too long; don't know / doesn't apply
Q40. Did you have confidence and trust in the out-of-hours clinician you saw or spoke to?	Yes, definitely; yes, to some extent; no, not at all; don't know / can't say	Did you have confidence and trust in the out-of-hours healthcare professional you consulted with?	No changes made
Q41. Overall, how would you describe your experience of the out-of-hours GP service?	Very good; fairly good; neither good nor poor; fairly poor; very poor	No changes made	No changes made

Changes made to the wording and response options of the four GPPS items evaluating out-of-hours care are underlined.

Description of questionnaire

The questionnaire comprised two sections. Section 1 contained the four modified GPPS evaluative stem items (applicable to all participants). These four items assessed service users' ratings of the 'entry access' to the service, the 'timeliness of care' received, their 'confidence and trust' in the health professional they consulted with, and their 'overall experience' of the out-of-hours service. Section 2 comprised the OPQ, which is composed of seven sections designed to capture information on the entirety of service users' experience of out-of-hours care. The composition of the OPQ has been detailed elsewhere, [8] where it was found to be both valid and reliable. Participants' ratings on fourteen evaluative items were analysed (Table 52); these were not management-specific and assessed users' experience of entry to the service, the outcome of their call, and the consultation with a health professional.

Table 52. The Out-of-hours Patient Questionnaire: 14 items used in analyses

Questionnaire section	Item	Response scale
Making contact with the service	How do you rate [how long it took your call to be answered, excluding any introductory message]?	5-point: ‘very poor’ to ‘excellent’
	Please rate the helpfulness of the call operator.	5-point: ‘very poor’ to ‘excellent’
	Please rate the extent to which you felt the call operator listened to you.	5-point: ‘very poor’ to ‘excellent’
Making contact with the service	How do you rate [how long it took for a health professional to call you back]?	5-point: ‘very poor’ to ‘excellent’
	Were you happy with the type of care you received?	Yes/no
	How do you rate [the length of your consultation with the health professional]?	5-point: ‘very poor’ to ‘excellent’
	[Please rate] the thoroughness of the consultation.	5-point: ‘very poor’ to ‘excellent’, plus N/A
Outcome of your call	[Please rate] the accuracy of the diagnosis.	5-point: ‘very poor’ to ‘excellent’, plus N/A
Consultation with the health professional	[Please rate] the treatment you were given.	5-point: ‘very poor’ to ‘excellent’, plus N/A
	[Please rate] the advice and information you were given.	5-point: ‘very poor’ to ‘excellent’, plus N/A
	[Please rate] the warmth of the health professional’s manner.	5-point: ‘very poor’ to ‘excellent’, plus N/A
	[Please rate] the extent to which you felt listened to.	5-point: ‘very poor’ to ‘excellent’, plus N/A
	[Please rate] the extent to which you felt things were explained to you.	5-point: ‘very poor’ to ‘excellent’, plus N/A
	[Please rate] the respect you were shown.	5-point: ‘very poor’ to ‘excellent’, plus N/A

Sampling

Sampling took place within two weeks of the service user contacting the out-of-hours service. The contact and demographic details for a random sample of 2,000 service users were extracted from electronic records at each site. Exclusion criteria were: age 12-17 years, due to the risk of breaching patient confidentiality on account of a questionnaire sent to the patient's home address, and because the GPPS targets those aged 18+; admission to hospital as a result of the contact; palliative care needs; or a temporary/incomplete address. After all exclusions were applied, a questionnaire, accompanied by covering letters from the research team and service provider, an information sheet and pre-paid envelope, was sent to a consecutive sample of the first eligible 850 service users (or parent or guardian if the service user was a child) from the sampling frame at each site. In one area, only 818 service users were sampled due to logistical constraints in the screening process. The total sample approached therefore totalled 5,068 service users. A reminder was sent two weeks after the initial mailing to non-respondents. Implicit consent was assumed if a completed questionnaire was received by the research team; no reminder was sent to service users who returned a blank questionnaire. Data collection took place between September 2013 and July 2014.

Data analysis

Respondents were compared to non-respondents in respect of their age, gender, deprivation quintile (using service users' postcodes to derive their Index of Multiple Deprivation 2010), and management option received as a result of the last recorded contact (from service provider record: telephone advice, treatment centre attendance, home visit) using a multi-level logistic regression model, clustering respondents by the provider from which they were sampled.

Construct validity

Construct validity of the four modified GPPS items was assessed by ascertaining how well they summarised the OPQ. First, a confirmatory factor analysis was conducted to establish whether the OPQ possessed the same two-factor structure reported in the paper detailing its development. [8] The standardised factor loadings with 95% confidence intervals for this model are reported. As Hu and Bentler [266] suggest, goodness of fit of the model was assessed through a two-index strategy using the Standardised Root Mean Squared Residual (SRMSR) supplemented with the Comparative Fit Index (CFI), [267] neither of which are adversely affected by large sample sizes. [268]

A principal component analysis (PCA) of the four modified GPPS items was then conducted to establish their latent structure, using the polychoric correlation matrix to account for the ordinal nature of these items. [269] Inspection of eigenvalues and component loadings were used to explore the underlying structure of responses. Based on this PCA, the construction of scales using the modified GPPS items and their internal consistency (Cronbach's α) was explored. Finally, the correlations between scales constructed above and the factor scores from the confirmatory factor analysis of the OPQ were investigated to assess the extent to which the modified GPPS item set summarised the OPQ.

Consultation satisfaction scale

The OPQ includes nine items rating service users' satisfaction with their consultation with an out-of-hours clinician (Table 52). These items were combined into a 'consultation satisfaction' scale, as suggested by the paper validating the OPQ, to avoid issues of multicollinearity in the regression models. To achieve this, each item was linearized to a 0-100 scale and respondents' mean scores from the nine items were derived as their 'consultation satisfaction' scale-score, provided they had answered at least four of the items. Finally, the scale was standardised so that the regression modelling would produce standardised coefficients.

Concurrent validity

To investigate the concurrent validity of the modified GPPS items, four multi-level linear regression models were constructed, creating a separate model for each evaluative outcome. The covariates were the management non-specific items from the OPQ (Table 48), including the 'consultation satisfaction scale'. Concurrent validity was considered to be established if each modified GPPS outcome was found to be significantly associated with thematically-related items from the OPQ. Univariate analyses were undertaken first, with covariates being excluded from the final models if they were not associated (p<0.10) to any of the four outcomes. All models controlled for service users' age, gender, deprivation quintile and management option, as well as the type of provider contacted (NHS, commercial, not-for-profit), and were clustered by provider. Missing data was accounted for using multiple imputations. To ensure the regression coefficients of covariates were comparable across models, the four modified GPPS outcomes, which originally had differing response scales (Table 51), were standardised. Sensitivity analyses were conducted to test for a linear trend over the covariate rating length of time taken for a health professional to call back, modelling the data whilst excluding those who were ‘not applicable’ (*n*=192). All analyses were performed using Stata v.13 (StataCorp, Texas, USA).

Results

Response rate and sample

Completed questionnaires were received from 1,396/5,068 (27.6%) of sampled service users. The multi-level logistic regression assessing response indicated that responders were older and more affluent (lower IMD score), but did not differ with respect to gender. Differences in response rates were also evident across the management options (Table 53). The response distributions for all variables of interest are displayed in Supplementary Table 1, in Appendix 42.

Table 53. Characteristics of responders and non-responders (n=5,067)

	Responders	**Non-responders**	***p*-value***
Frequency (%)	1,396 (27.6)	3,672 (72.4)	
Age in years, mean (SD)	46.0 (28.2)	32.5 (26.2)	<0.001
Gender, female (%)	877 (62.8)	2,208 (71.6)	0.081
IMD score, mean (SD)	19.0 (14.0)	23.9 (15.9)	<0.001
Management option			
Telephone advice (%)	492 (35.2)	1,143 (38.5)	0.001
Treatment centre (%)	647 (46.4)	1,765 (48.1)	
Home visit (%)	172 (12.3)	301 (8.2)	
Other (%)	85 (6.1)	193 (5.3)	

Reported *p*-values were obtained from a multi-level logistic regression that compared responders to non-responders. The model clustered individuals by the provider from which they were sampled.

Construct validity

Confirmatory factor analysis of the Out-of-hours Patient Questionnaire

The confirmatory factor analysis revealed that the data fit the proposed ‘entry access’ and ‘consultation satisfaction’ two-factor structure reported by Campbell et al. [8] moderately well (Table 54), with a SRMSR of 0.06 (values under 0.08 represent good fit) and a CFI of 0.89, which is just short of the suggested cut-off of 0.90 for good fit. [266] In-line with Campbell et al., the two latent variables were moderately correlated (r=0.54, *p*<0.001). [8]

Table 54. Confirmatory factor analysis of the Out-of-hours Patient Questionnaire

	Coef.[1]	**95% CI**	***p*-value**
Entry Access			
How do you rate [how long it took your call to be answered]?	0.65	0.61-0.70	<0.001
Please rate the helpfulness of the call operator.	0.91	0.89-0.93	<0.001
Please rate the extent to which you felt the call operator listened to you.	0.90	0.88-0.92	<0.001
How do you rate [how long it took for a health professional to call you back]?	0.66	0.62-0.70	<0.001
Consultation Satisfaction			
Were you happy with the type of care you received? [no/yes]	0.47	0.41-0.52	<0.001
How do you rate [the length of your consultation with the health professional]?	0.80	0.77-0.83	<0.001
[Please rate] the thoroughness of the consultation.	0.88	0.86-0.89	<0.001
[Please rate] the accuracy of the diagnosis.	0.84	0.81-0.86	<0.001
[Please rate] the treatment you were given.	0.86	0.84-0.88	<0.001
[Please rate] the advice and information you were given.	0.90	0.88-0.91	<0.001
[Please rate] the warmth of the health professional’s manner.	0.87	0.85-0.89	<0.001
[Please rate] the extent to which you felt listened to.	0.93	0.92-0.94	<0.001

[1] Item loading

Principal component analysis of the modified GPPS items

The PCA of the four modified GPPS items extracted a single component with an eigenvalue exceeding 1.0 (2.78), which accounted for 69.5% of the variance in the data. Observed component loadings were 0.44 for ‘entry access’, 0.47 for ‘timeliness of care’, 0.51 for ‘confidence and trust’ and 0.57 for ‘overall experience’. This component can be interpreted as overall satisfaction with out-of-hours care. A rotation was unnecessary, as simple structure was obtained.

Informed by the PCA, we investigated the construction of an ‘overall satisfaction’ scale using all four items. This scale was derived by summing the standardised items (to account for differing response scales), if responses were given to all items. The scale had acceptable internal consistency, $\alpha=0.772$. Excluding the ‘entry access’ item suggested a very minor improvement in alpha, $\alpha=0.777$ (Supplementary Table 2, Appendix 41).

How well do the modified GPPS items summarise the OPQ?

The ‘overall satisfaction’ scale was highly correlated with the factor scores of both OPQ domains for ‘entry access’ ($r=0.63$, $p<0.001$, $r^2=0.397$) and ‘consultation satisfaction’ ($r=0.66$, $p<0.001$, $r^2=0.440$). These correlations are both stronger than the correlation reported between the two OPQ domains. When combined into a scale, the four modified GPPS items explain 39.7% of the variation in ‘entry access’ factor-scores and 44.0% of the variation in ‘consultation satisfaction’ factor-scores, summarising both scales moderately well. Table 54 reveals that the ‘entry access’ domain of the OPQ was most related to service users’ experience of the call-operator, for which there is no equivalent GPPS item, perhaps explaining the lower correlation between the ‘overall satisfaction’ scale and the ‘entry access’ factor-scores.

Concurrent validity

Multiple imputation of missing data allowed for inclusion of all 1,396 respondents in the four mixed-effects multi-level linear regressions. A divergent pattern of associations across the covariates was evident between the models for each of the four GPPS outcomes (Table 55).

Covariate	Entry access		Timeliness of care		Confidence and trust		Overall experience	
	Coef. (95% CI)	*p*	Coef. (95% CI)	*p*	Coef. (95% CI)	*p*	Coef. (95% CI)	*p*
Call answer time	0.13 (0.06, 0.21)	0.001	0.09 (0.03, 0.15)	0.006	0.00 (-0.06, 0.05)	0.945	0.01 (-0.05, 0.07)	0.808
Helpfulness of operator	0.14 (0.04, 0.24)	0.008	0.06 (-0.03, 0.15)	0.204	0.04 (-0.04, 0.12)	0.345	0.12 (0.04, 0.20)	0.003
How operator listened	0.15 (0.05, 0.25)	0.003	0.05 (-0.04, 0.14)	0.268	0.00 (-0.08, 0.09)	0.954	0.07 (-0.01, 0.15)	0.068
Health professional call back time[1]	0.09 (0.03, 0.16)	0.007	0.45 (0.39, 0.52)	<0.001	0.05 (-0.02, 0.11)	0.140	0.13 (0.08, 0.19)	<0.001
Very poor / poor	Reference group		Reference group		Reference group		Reference group	
Acceptable	0.16 (-0.02, 0.34)	0.089	0.70 (0.54, 0.86)	<0.001	0.07 (-0.09, 0.23)	0.376	0.38 (0.24, 0.52)	<0.001
Good	0.34 (0.15, 0.53)	0.001	1.05 (0.87, 1.22)	<0.001	0.18 (0.02, 0.35)	0.030	0.51 (0.37, 0.66)	<0.001
Excellent	0.35 (0.14, 0.56)	0.001	1.41 (1.22, 1.60)	<0.001	0.10 (-0.08, 0.29)	0.271	0.48 (0.31, 0.64)	<0.001
Not applicable	0.29 (0.07, 0.52)	0.011	0.98 (0.79, 1.17)	<0.001	-0.04 (-0.23, 0.15)	0.706	0.35 (0.17, 0.53)	<0.001
Happy with treatment option								
Yes	Reference group		Reference group		Reference group		Reference group	
No	-0.21 (-0.39, -0.02)	0.030	-0.32 (-0.49, -0.15)	<0.001	-0.58 (-0.73, -0.44)	<0.001	-0.70 (-0.83, -0.56)	<0.001
Consultation satisfaction	0.05 (-0.01, 0.12)	0.105	0.06 (0.01, 0.12)	0.025	0.56 (0.51, 0.61)	<0.001	0.43 (0.38, 0.47)	<0.001

[1] Sensitivity analyses excluded the 'not applicable' category and entered this covariate as an ordinal variable to assess the global effect on each outcome, which is reported above the effects of the separate dummy variables. Models controlled for participants' age, gender, deprivation quintile (IMD), ethnicity (White, other ethnic group), and management option received (telephone advice, treatment centre, home visit/other), as well as the type of provider contacted (NHS, commercial, social enterprise), and were clustered by provider (n=6). The GPPS items (dependent variables) were standardised so that regression coefficients are comparable across models. For all models, n=1,396.

Table 55. Linear regression models showing the associations of OPQ items to the four modified GPPS outcomes

Discussion

This study sought to determine the construct and concurrent validity of four items from the GPPS [119] evaluating service users' experience of out-of-hours care through comparisons to an established, valid and reliable measure; the OPQ. [8, 256] Preliminary work highlighted the need to make minor modifications to three of the four GPPS items to improve comprehension by service users', and response options. The modified GPPS item-set ('entry access', 'timeliness of care', 'confidence and trust', 'overall experience') formed a single scale, which summarised the two-domain structure of the OPQ moderately well. Therefore, given minor modifications, these findings indicate that the GPPS item-set evaluating out-of-hours care has potential for acceptable construct validity as a scale of overall satisfaction.

Each of the four outcomes was strongly associated with a distinct set of thematically-related items from the OPQ, demonstrating their concurrent validity. Evaluations of entry access were related to ratings of the length of time before service users' calls were answered, the helpfulness of the call operator, and the extent to which the operator listened, which is supported by these items loading onto the same construct in PCAs in the present study and elsewhere. [8, 119] Similarly, evaluations of timeliness of care were significantly associated with time taken for the call to be answered, but were not related to ratings of the helpfulness of the call operator. Instead, timeliness was most strongly associated with the length of time taken for a call back from a health professional, an association also observed in a recent study of patient satisfaction with out-of-hours care from the Netherlands. [264]

Croker and Campbell [139] found that patients' confidence and trust in a health professional with whom they consulted in an in-hours primary care setting was highly influenced by interpersonal aspects of the care delivered as reported by patients. Important characteristics included having been given enough time, having felt listened to, having been given explanations about tests and treatments, having treated the patient with care and concern, and having taken them seriously. In the present study, analogous items from the OPQ, combined into the consultation satisfaction scale, were strongly associated with service users' ratings of confidence and trust in the out-of-hours health professional they consulted with. Confidence and trust was not related to items evaluating entry access.

Respondents' ratings of their overall experience were strongly related with items from all three included sections of the OPQ; entry access, the result of the user's call and the consultation with a health professional. The consultation satisfaction scale included an item rating the length of the consultation, which has also been shown to be a factor in confidence and trust. [270] Patients' evaluations of their overall experience of in-hours primary care have been shown to be most associated with doctor communication and the helpfulness of receptionists. [94] In the present study, service users' ratings of their overall experience (the item unmodified from the GPPS) were strongly associated with their consultation satisfaction, which included elements of doctor communication, as well as the helpfulness of the call operator.

Strengths and limitations

A strength of this study is the large sample of service users, which facilitated reliable statistical analyses using a large number of variables. When using factor analysis, best practice is to have 5-10 participants per measure, [268] with a higher participant-to-measure ratio yielding more reliable results; upwards of 64 participants per measure were used in these analyses.

The overall response rate was low and responders tended to be older and living in less deprived areas, and the final respondent sample had a higher proportion of males in non-responders. This threat to the representativeness of the study sample is unlikely to have affected the analyses reported here. Specifically, this analysis aimed to determine the structure of users' experience items and on associations between them, rather than providing incidence/prevalence rates of conditions or similar outcomes that are might be more affected by response bias issues. The methods employed controlled for these factors where possible and the findings are corroborated by the existing literature, as discussed above.

Minor modifications to either the word stems or to response categories for three of the four GPPS items were made after careful piloting with service users that included the use of cognitive testing. Furthermore, the GPPS asks questions to respondents about making contact with a GP out-of-hours service in the past six months, whilst this study's respondents were asked to answer questions relating to the last time they made contact with a GP out-of-hours service, having been sampled from out-of-hours providers' databases within two weeks of having made contact. While this may limit the degree to which these findings apply to the existing GPPS survey items somewhat, this piloting was essential as early feedback from service users identified problems interpreting the items and changes to two items were designed to minimise missing data through blank responses (e.g. missing response categories). Implications for practice based on these findings are therefore contingent on the adjustment of current GPPS items.

Workstream 3. Exploring how out-of-hours services use patient feedback

Study aims and objectives

This study aimed to identify how out-of-hours GP providers routinely collect patient experience feedback (including GPPS data) to inform their practice, with a particular focus on how it can be used to inform service reconfiguration and to improve patient experiences of out-of-hours care. This was achieved by undertaking qualitative interviews with staff from out-of-hours service providers.

Methods

Sampling and data collection

The aim was to recruit an additional six out-of-hours providers as six (n=12) were already recruited and had taken part in the survey study (see report on the conduct of Workstream 2, above). Provider and staff recruitment ceased when data saturation was achieved. To achieve diversity of high, medium and low scoring services, providers were first sampled on the basis of their scores for the GPPS item for care received from the service (Question 40, April–September 2010 national GPPS dataset). Once categorised into these groupings, information on organisation type and geographical location were considered. The final sample of providers ensured diversity across these three domains (GPPS score, organisation type and location), although no comparison of different sub-groups of providers was planned. Up to three potential interviewees who had some involvement in conducting patient experience surveys were identified and approached to be interviewed in each provider. Participants were provided an information pack consisting of a covering letter and participant information sheet. A mutually convenient time was organised to conduct the interview.

A week before the interview participants were sent a copy of a ‘feedback report’ containing patient ratings of their provider organisation based on the July 2012-March 2013 wave of GPPS. Benchmarking data (generated by matching GP practice postcodes to provider localities) were produced to allow providers to compare their performance to that of the 91 other English out-of-hours services for whom scores were able to be generated. Reports for the six services that had participated in the survey study (Workstream 2) also included a summary of the provider's ratings derived from the research survey.

Face-to-face interviews, conducted at the participant's workplace, took place between April and July 2014, each lasting between 39 and 88 minutes (mean: 59 minutes). Topic guides were developed from a literature review, discussion between researchers and providers and previous findings with comments provided by the study advisory group. The topic guide included questions on how providers collected patient experience data and how this was used to make service changes; awareness and views of GPPS and out-of-hours items within it; reflections on the use of GPPS benchmarking provided in their feedback report.

Analysis

Interviews were digitally recorded, transcribed verbatim and transcripts were checked against the original recording for accuracy. Transcripts were coded in NVivo v.10 software (QSR International Ltd, 2012), and analysis was independently coded using an iterative approach by one researcher (HB). A sample of five transcripts were independently analysed by a second coder (AA) to ensure agreement was reached on the coding frame and codes. A deductive, framework approach with preliminary codes reflecting the content of topic guides was used to construct the coding framework. However, a more inductive approach with additional thematic coding was undertaken using the 'constant comparison' method [270] to capture new themes emerging from the dataset. The initial coding frame was discussed within the team, and where possible the codes were tested through seeking negative cases and/or divergent data. The data were then reorganised and collapsed into over-arching themes. This process took place on two occasions, until the main categories were agreed. All participants were sent a summary of findings with a structured feedback form inviting comments on the veracity of the interpretation of the study findings. Final themes were reviewed and agreed between the research team to enhance reliability.

Results

Study participants

Five of the six providers approached took part (in addition to the six who participated in the survey study). A total of 31 staff from the 11 providers (n=2 NHS organisations, n=4 social enterprises, n=5 commercial organisations) were interviewed, at which point data saturation was judged to have been achieved. Most participants were female (n=23); 18 were service managers, seven were clinicians (GPs), and six were administrators. Participants who completed the feedback form (n=2) on the findings were satisfied with the accuracy of the summary. Three main themes emerged: using surveys as a method of obtaining patient feedback; the utility of patient feedback; and the value of benchmarking.

Surveys as the most common method of obtaining patient feedback

Most participants focused on survey methods of collecting patient feedback, as ten of the eleven providers undertook regular surveys to audit their patients' experiences. Participants also discussed the ambiguities of operationalising NQR5, the desire for qualitative feedback to supplement survey data, and the role of alternative methods in addition to surveys. It was evident from discussions that each provider interpreted the sampling for NQR5 differently; for example, the range of patients being routinely audited varied from 1% to 20%:

> *'We send out approximately 250 a week. Our National Quality Requirements require us to survey 1% - we actually do considerably more than that because we have taken our own interpretation on it…'*
>
> *(11_4001, Manager)*

Audits were undertaken on either a weekly or monthly basis, using survey instruments developed by the organisation. Some participants reported that weekly audits were useful in terms of maximising patient response rates:

> ‘*…they’ve [out-of-hours service] worked out that the sooner the patient gets the questionnaire the more likely it is that they will complete it because it's still fresh in their minds, so they try to do it as quickly as possible*.’
>
> *(14_4003, GP)*

Most participants placed great importance on qualitative feedback from free-text comments provided by patients which helped to interpret the quantitative findings, identify actions and provide a more personalised response from patients:

> ‘*If they have got a real issue they can put it down, can’t they? Just doing the survey itself is just a way you test the water…. The free-text allows someone who has got a very bad experience the opportunity to write to us.*’
>
> *(10_4001, Manager)*

> ‘*I’m dealing with people, I’m not dealing with robots. I mean, it’s their experiences, their feelings and they need to have a place to feed that back… they absolutely need to have a place to express their opinions – that’s giving people a voice.’*
>
> *(14_4003, GP)*

Although patient surveys were agreed to as a necessity by all but one of the participating providers, this was not a sufficient resource to drive change within services. A wide variety of alternative methods used by providers were reported, such as comment cards, ‘complaint and compliment systems’ and new technologies:

> *‘At the moment we’re thinking of going more electronically, so as soon as you have your consultation in the base, you come out and there’s a tablet so you can actually do your surveys straight after… that way you can get more accurate feedback of how people are feeling*.’
>
> (*19_4002, Administrator)*

Utility of patient feedback

Many participants cited examples of ways in which patients' reported experiences had been used to make changes to service provision, although most changes tended to be 'low-level'. Due to the lack of observed trends within the data most participants reported that patient survey data was insufficient to instigate service-wide changes:

> *'In the main the results are stable and pretty good, but there's not enough that's consistent that I think we could use around wholesale service change.'*
>
> *(12_4003, Manager)*

Participants reported that patients' expectations of the out-of-hours service were often unrealistic and difficult to manage, and this made patient feedback difficult to deal with:

> *'You often get patients who are very unhappy about the service they got and when you drill down into it it's because they didn't get antibiotics for their cold. Its expectations.'*
>
> *(16_4003, GP)*

The changing landscape of the urgent care system was also confusing to patients. Some staff participants questioned the validity of patient experience data as the patients may be unaware of the different elements of the care pathway. Another barrier identified was the low level engagement by commissioners. Despite the fact that patient experience audits are part of NQR5, many participants reported that commissioners treated it as a 'tick-box' exercise:

> *'They [the commissioners] don't come across to me as particularly engaged in this at all, and never really ask us too many questions around it.'*
>
> *(18_4003, Manager)*

Although acknowledging the identified barriers, some participants discussed how engaging with patient feedback had subtly changed the culture within their organisation, and highlighted the importance of transparency and being responsive to change. In addition, participants reported the benefits of being able to compare patient feedback with other areas of reporting within the NQRs.

Value of benchmarking

Most participants acknowledged the benefits of having access to benchmarking data and felt it was a facilitator to enabling change. Notwithstanding this, many staff interviewees placed greater importance on their own surveys over the GPPS data, largely as their own surveys were more detailed.

Some staff expressed concerns about the reluctance of some providers to share with and learn from other providers, an issue mainly arising from commercialisation taking place within the NHS:

> ‘*It’s terrible isn’t it, when everybody’s competing and not collaborating? That’s the system we’re living with, we’ve had to get used to it.*’
>
> *(18_4001, GP)*

The benchmarking provided using the GPPS out-of-hours patient ratings were seen as useful, although many identified weaknesses with set items as they felt that the questions did not reflect the current urgent care system and lacked detail:

> ‘*It is [General Practice Patient Survey out-of-hours evaluative items] just four questions, you get asked in McDonalds. It’s not detail is it?*’
>
> *(10_4001, Manager)*

Discussion

In the UK out-of-hours primary care providers are mandated to regularly audit patients' experiences as part of the NQRs, and services routinely met this requirement by conducting patient surveys, as well as by obtaining feedback using a variety of other methods. However, NQR5 is ambiguous and the resultant data cannot be used to compare services as providers are undertaking audits of varying scale, frequency, and methodology. Staff reported a strong preference for qualitative patient feedback, which is echoed in other settings, as it yields richer, more detailed feedback than quantitative survey scores. For example, hospital staff have found qualitative data from patients added a more patient-centred aspect to patient satisfaction measurements. [195, 272] Research has shown that healthcare leaders placed great importance on complaints, comments and compliments as a source of patient feedback, [273] as did GP practice staff (Chapter 7).

Patient feedback appeared to have a limited role as a driver for service change and effective change was hindered by modifications taking place in the urgent care landscape, which confused patients as to how care was organised. Some staff also reported that commissioners appeared uninterested in patient experience audit findings. In some settings audit and feedback has been shown to have small to moderate effects on healthcare professionals' practice, [189, 274] although in other settings it can have a wider impact. [275] In order for change to occur, the organisational culture must be supportive of change and be patient-focused. [191, 194, 276] Most of the changes reported by staff were 'low-level' and unlikely to drive system-wide reconfiguration due to the lack of consistent patterns observed in the data. There was a preference for qualitative feedback as patient free-text comments could potentially identify specific areas of actionable change, or contribute to wider data gathering audits, e.g. critical incident techniques. [277] However, in order to be useful; patients' attention must be focused to provide qualitative feedback on the out-of-hours service.

Staff valued the GPPS patient experience benchmarking data and the GPPS presents an opportunity for benchmarking of all out-of-hours services. NHS England has recently recommended that NHS commissioners use the GPPS results to monitor patient experiences of out-of-hours providers, [252] and the CQC has published GPPS provider performance at commissioner level. [251] Despite the strengths of the GPPS (regularly and independently collected data which is publically available), participants were reluctant to use GPPS data in its present form due to concerns about the face validity of out-of-hours items and the absence of free-text comments, a limitation found in previous studies. [111, 278] In addition, the current out-of-hours items are not reflective of the recent changes that have taken place within the urgent care system (e.g. introduction of the NHS '111' telephone portal). Most staff did not believe that the limited number of GPPS items would drive change by themselves.

Strengths and limitations

This is the first qualitative study to explore the views of out-of-hours staff who have an in-depth knowledge of patient feedback processes within their organisation. Sampling ensured that staff from a variety of different types of provider (e.g. not-for-profit, or commercial enterprises), serving diverse populations across England, were included. Although sampling diversity was achieved, it is acknowledged that participating organisations may be more interested in the patient experience agenda than non-participants, and thus findings may not reflect the views of the wider population. The views of commissioners were not sought in this study, thus the widespread perception that some commissioners were apathetic towards patient feedback data must be interpreted cautiously. Due to logistical constraints it was not possible to interview commissioners and obtain their perspective on their perceived role and value.

Conclusions from the out-of-hours research

Implications for practice and future research

An analysis of national GPPS data (Workstream 1) identified that commercial providers were associated with poorer patient experiences of out-of-hours GP care compared with NHS or not-for-profit providers. Simple explanations regarding the drivers of these lower ratings are not possible in this observational dataset, and further research is required to understand what is driving these differences. While some insight might be gained from an understanding of patient differences (e.g. nature or urgency of request for care) at the level of the provider, such data are not routinely collected in GPPS for out-of-hours service evaluations. It is unknown whether factors, such as user awareness of the provider type, may also be of importance in interpreting service users' ratings.

Future research, possibly involving qualitative approaches or a vignette study, is required to investigate reasons for generally lower scores from service users from minority ethnic backgrounds (See Chapter 6 for vignette work conducted as part of the wider IMPROVE programme). Similarly research investigating reasons why service users who were unable to take time from work to attend their practice during regular hours reported poorer scores across all three evaluative questions is needed. Finally, as for in-hours GP care, [209] investigation of the extent to which variations between socio-demographic groups in respect of care ratings might be attributable to the clustering of servicer users belonging to socio-demographic groups reporting relatively lower scores within providers with lower overall scores is required. This analysis would help inform the development and targeting of interventions aimed at improving service users' experience of out-of-hours GP care for specific population sub-groups.

National standards (NQR5) require out-of-hours providers to routinely audit patient experiences, although no specific survey tools or methods are recommended to achieve compliance. In the absence of data collected directly by providers, both the National Audit Office and the CQC have recently used GPPS as an alternative data-source to monitor patient experiences of GP out-of-hours care. An important prerequisite to using GPPS data to benchmark services is, however, that its psychometric properties are established. The reliability of GPPS out-of-hours items have been previously reported, [133] but there was no evidence regarding its validity. The second workstream demonstrated that whilst our survey composed of only four of the GPPS evaluative items (after minor but essential modifications identified through cognitive testing and piloting), the GPPS out-of-hours items we used had both construct and concurrent validity. These findings provide support for the use of GPPS for national benchmarking purposes.

While Workstreams 1 and 2 examined the technical performance of GPPS out-of-hours items, the third workstream examined how out-of-hours staff use patient feedback, and their views towards the utility of GPPS items. This qualitative study found that while national quality requirements (NQRs) are intended to promote transparency and allow comparisons between out-of-hours providers, NQR5 was ambiguous and in its current form does not support benchmarking or service improvement. A critical review of the NQRs is required to help providers to engage with patient feedback and drive service improvement effectively.

In the absence of clear NQR guidance, providers were inventive in the ways in which they engage with patients. Qualitative feedback was highly valued as it provided detailed information which could lead to actionable changes. However, services struggled to find ways to use patient feedback to drive anything other than low-level service change. Future research should explore how out-of-hours services managing patients with urgent care needs, and particularly those delivering services to diverse populations, can be assisted in engaging more fully with patient feedback. Evidence is also needed on whether comprehensive guidance on how to collect, interpret and act upon patient feedback has the potential to drive quality improvement initiatives. [42, 191, 276]

In the context of the rapidly changing landscape of UK urgent care services, although participating providers could see the potential of using GPPS for benchmarking purposes, its out-of-hours items need urgent revision as they do not reflect current telephone access arrangements (NHS 111) for out-of-hours care. This qualitative finding supports our preliminary survey piloting work and cognitive interviews with service users (Workstream 2). Minor, but essential amendments to GPPS out-of-hours items are required to improve the comprehension of items and improve data quality.

Patient feedback currently has a limited role in driving changes to out-of-hours service provision, and the utility of feedback may be hindered, in part, by recent modifications to the urgent care system and the ambiguity of NQR5 relating to gathering and acting upon patient feedback. English GPPS data may be used to benchmark and compare service providers. However, the out-of-hours items need to be updated to reflect the changes made to accessing out-of-hours services by telephone, so that providers can be confident that ratings reflect their services' performance. A greater understanding of how variations in patient and provider characteristics drive variations in patient experiences of out-of-hours care is needed to support the development and targeting quality improvement initiatives.

Chapter 12. Conclusions, implications for practice, and recommendations for future research

Conclusions

In chapter 1 we outlined how, following the introduction of a wide range of quality improvement strategies as part of an overarching ‘clinical governance’ strategy in the late 1990s, there had been step changes in the management of the major chronic diseases in the NHS. However, the ways in which patients experienced healthcare had not been given such a priority, and the need for a rebalancing was seen by increasing attention to patient experience in policy documents, routine publication of patient experience data, benchmarking of hospitals in relation to patient experience, and even an (ill-fated) attempt to attach payments to patients assessment of their GP’s care.

There has therefore been widespread acceptance that good patient experience is an important outcome of care in its own right, and our work [279] and that of others [280] have shown that patient experience is a domain of quality that is distinct from but complementary to the quality of clinical care. While there have been an increasing number of surveys developed to measure patient experience, there has been equally widespread acceptance that these measures have not been very effective in actually improving care. [42] This is the background to our programme of work. Entitled ‘IMPROVE’, we aimed to find better ways of both measuring and using information on patient experience that would lead to improvements in patient care in both in-hours and out-of-hours primary care settings.

In the introduction, we described a range of ways of getting patient feedback on their care – including surveys, focus groups, and analysis of complaints. In this programme, we have focused on the use of patient surveys as they are the dominant method currently used in the UK. However, in chapter 10 we describe an exploratory trial of real time feedback, which moves away from the paper based questionnaires which still dominate the measurement of patient experience in the NHS.

This programme had seven aims, each of which was tied closely to one work-package of research. These aims were:

1. To understand how general practices respond to low patient survey scores, testing a range of approaches that could be used to improve patients’ experience of care.
2. To estimate the extent to which aggregation of scores to practice level in the national study masks differences between individual doctors.
3. To investigate how patients’ ratings on questions in the GP Patient Survey relate to actual behaviour by GPs in consultations.
4. To understand better patients’ responses to questions on communication and seeing a doctor of their choice.
5. To understand the reasons why minority ethnic groups, especially South Asians, give lower scores on patient surveys compared to the White British population.
6. To carry out an exploratory randomised controlled trial of an intervention to improve patient experience, using tools developed in earlier parts of the programme.
7. To investigate how the results of the GP Patient Survey can be used to improve patients’ experience of out-of-hours care.

The aims of the programme have not changed during the five years of our research, though some details of the research have been modified as the work progressed (we have summarised any changes in each individual chapter). We presented the results of our research under three broad headings, and also use these headings for this final chapter of discussion and conclusions, namely:

- Understanding patient experience data (aims 3 and 4)
- Understanding patient experience in minority ethnic groups (aim 5)
- Using data on patient experience for quality improvement (aims 1, 2, 6 and 7)

1. Understanding patient experience data

Patient surveys are now widely used in many countries, yet still comparatively little is known about what experiences lead patients to respond in particular ways in these surveys. What drives them to tick particular boxes, and how do those responses relate to the care they have actually received? We approached this in two main studies, one in which we asked patients directly about how they chose certain items on the questionnaire while showing them a video of their consultation (chapter 2) and one in which we compared their responses with those of expert raters using two standard instruments for assessing videos of consultations (chapter 3). The results of these studies have important implications for the interpretation of survey data, particularly that focussed on patient evaluations of specific encounters with health care professionals.

The first study (chapter 2) showed that while patients readily criticised their care when reviewing GP consultations on video, they described how they had been reluctant to be critical when completing a questionnaire after the consultations. Reasons for this included the need to maintain a relationship with the GP (including uncertainty about how confidential survey results would be), and their gratitude for the care they had received from the NHS in the past. In addition, perceived power asymmetries made people reluctant to criticise their doctor. Patients were also disinclined to be critical when completing a questionnaire if they had actually received the treatment they wanted. Overall, we concluded that patients find questionnaires administered at the point of care may be limited tools for being able to feedback concerns about primary care consultations.

The second study (chapter 3) reinforced our conclusion from chapter 2 that patient evaluations of consultations in surveys may present an uncritical view of the actual consultation. In this study we had videotapes of GP-patient consultations assessed by four independent clinical raters. The results were striking. When trained raters rated communication within a consultation to be of a high standard, patients did the same (with one single exception). However, when trained raters judged communication within a consultation to be of a poor standard, patients’ assessments varied from poor to very good. This finding again points to the reluctance of patients to criticise their doctors in questionnaire surveys. In the previous study the ‘gold standard’ was the patient’s own account of the consultation, and in this study the standard was that of a trained external GP rater.

We do not think these results mean that patient surveys can’t be used to assess the quality of general practice care. However, they do point to clear limitations. One of the concerns that GPs have about surveys (see chapters 7 and 8) is that surveys are selectively completed by critical or grumpy patients and that survey results will therefore give a negative and biased view of the GP’s care. The results of these two studies suggest that the opposite is the case. Patients’ reluctance to criticise their doctors means that survey responses using evaluative type of questions are likely to give an over-positive view of the doctor’s care. This is one reason why there has been a move towards using report items in some survey instruments (though we don’t know whether these suffer from similar problems). Because of this tendency for patients to choose the most positive response options, we suggest that absolute scores should be treated with some caution: they may present an over-optimistic view of the GP’s care. However, this does not mean that surveys cannot be used to look at relative scores: scores from a GP which are lower than their colleagues and from GPs in other practices are likely to indicate a problem, even though high scores from other doctors or practices may conceal deficiencies in care in those practices too.

We also looked at how GPs rated their own consultations. They completed a form immediately after each consultation, using the same scale as the patient. GPs were certainly more inclined to criticise themselves than the patients were to criticise the care they had received. This is entirely consistent with the findings from our subsequent interviews with patients. However, we found absolutely no correlation between patient and GP scores. Neither did we find any correlation between GP’s own scores and those of expert raters who reviewed the consultation on video. GPs are clearly using different parameters when assessing their own performance, but we were not able to investigate this in more detail in this study.

When we spoke to GPs about their survey results (chapters 7 and 8), through both focus groups and face-to-face interviews, they reported how, whilst positive about the concept of patient feedback, they struggled to engage with and make changes under the current approaches to measurement. They also commonly expressed concern that patients would be critical of their care if they didn’t get what they wanted (e.g. an antibiotic prescription). This concern was borne out to some extent by our results. In our analysis of the assessment of nurses (chapter 4), a strong predictor of survey scores was whether the patient wanted to see a nurse when they first contacted the practice. If they had wanted to see a GP but saw a nurse, then the scores given to that nurse were much lower. We have no reason to think that the nurse communicated worse in those consultations and the low scores may therefore indicate a more general dissatisfaction by patients in not having their original expectations met.

It is important to understand that, in line with the overall aims of the programme, the work in these two chapters focused on assessment of communication in the primary care consultation (such as giving the patient enough time, and explaining tests and treatments). Our conclusion that survey scores have more value in assessing relative performance than absolute performance of doctors may or may hold true for other aspects of practice performance commonly assessed in surveys, such as difficulty in getting appointments, getting through on the phone and waiting times. Patient’s reasons for not wanting to criticise their doctor may be less important when they assess what they regard as management aspects of the practice.

A second aspect of care which we identified as part of our programme of work relates to patients’ ability to see a GP of their choice. While most of our research focused on communication, the results we report in chapter 4 have some important findings in relation to patient choice. [134] The results first show that most patients have a particular GP whom they prefer to see. It is sometimes suggested that this only matters for some population groups (e.g. not for young people) but we find this is not the case. Even among 18 to 24 year olds, more than 50% of respondents to the GP Patient Survey have a particular doctor they prefer to see, rising to over 80% in people over 75. Disturbingly, a large percentage of people who have such a preference are unable to see the doctor of their choice. This percentage has risen from 30% to 40% over just the past five years. One possible impact of this change comes from our analysis of data from patients who have seen a nurse where patients had originally wanted to see a doctor expressed considerable dissatisfaction with their subsequent consultation with a nurse. However, these data do not reflect what would have happened if patients had seen another doctor, just not the one of their choice.

Overall, patients express more negative opinions about choice of doctor than in any other part of the GP Patient Survey, something that has may in part have got worse as a result of government policies to improve access. There is a clear tension between the ability of practice to provide rapid access and continuity of care, and data from our studies suggest that patients' inability to see a doctor of their choice is a significant quality issue for the NHS.

2. Understanding patient experience in minority ethnic groups

In this part of our research, we focused our main work on survey responses from minority ethnic groups and on South Asians in particular. The general interest in minority ethnic groups is because they tend to report worse experiences in surveys in most countries studied, as well as in the UK. Our research on out-of-hours care in this programme (chapter 11) replicated this result, with Asian and Mixed ethnic groups reporting worse experiences than the White majority.

Our specific focus in the major strand of this research was on South Asian respondents, because of the size of this group in England and the consistently low scores generated by this group in English surveys across both primary and secondary care settings. We focused on questionnaires competed in English: although the GP Patient Survey is available in 15 languages, a tiny minority of surveys are completed in languages other than English (typically <0.2% of returns).

A number of potential explanations have been suggested for the lower ratings given by South Asian and other minority ethnic groups. Broadly, these relate to whether South Asian patients (a) receive lower quality care, or (b) receive the same care, but rate this more negatively. [68] For example, South Asian respondents might rate the same care more negatively if they have higher expectations, or because they interpret the survey items and response options in different ways (such as being culturally less likely to check extreme options).

The last of these options was potentially the simplest to explore. Taking advantage of the large numbers available in the GP Patient Survey to examine the responses of South Asian groups using item response theory and allowing for a wide range of other socio-demographic characteristics (chapter 4, workstream 3), we found no evidence that South Asians used the scales in a different way to White British respondents. While these results do not provide conclusive evidence of equivalence in the way in which different respondents use the survey scales, they increase the likelihood that the worse experience reported by South Asians reflects either differences in expectations or genuinely worse care. Our previous work [68] suggested that for one aspect of care (waiting times), South Asians might have higher expectations of care, implying that their lower scores on surveys might not be associated with worse care. We were able to advance our understanding of this complex issue considerably as a result of the research in this programme.

First we showed that Asian respondents to the GP Patient Survey tend to be registered in practices with generally low scores. This explained about half of the difference in reported experience between South Asian and White British patients (chapter 5, workstream 1) and identified that some practice effects were related to the ethnicity of the doctor (minority ethnic doctors receiving lower scores for doctor patient communication (chapter 5, workstream 4). However, these practice effects did not account for the low scores among South Asian patients, even though the differences were reduced where the practice offered consultations in a South Asian language [158], (PhD project allied to our programme). Next we showed that, far from being uniform across all population groups, the lower scores from South Asian patients were much more marked among older female respondents. It was therefore important in our subsequent work to ensure that these patients were represented in our research (chapter 5, workstream 2).

In video elicitation interviews with South Asian patients (chapter 2), we identified the same issues driving evaluations of communication in South Asian and White British patients: their relationship with their GP (and others within the practice), their expectations of the consultation, and a reluctance to criticise doctor’s performance. The finding that South Asians are assessing broadly similar issues when completing questionnaires therefore still leaves unanswered the question of why scores from South Asian patients are low.

The final and most original part of this work provides insight into this (chapter 6). Here we filmed 16 simulated consultations based on transcripts of real consultations using various combinations of White and Asian doctors and patients, half scripted to be ‘good’ and half ‘poor’. We showed three randomly sampled videos to each of 1,120 people (half White British, half Pakistani, equally split between those under and over 55) and asked them to score the consultation using the communication items from the GP Patient Survey.

If the low scores reported by South Asians in real life settings were due to higher expectations on their part, then we would expect them to give lower scores in the experimental vignette situation. Quite the reverse happened. When viewing the same consultation, South Asian respondents gave scores which were higher – indeed much higher when adjusted for socio-demographic characteristics – compared to White British respondents. This suggests that the low scores given by South Asian patients in surveys such as the GP Patient Survey reflect care which is genuinely worse, and possibly much worse, than that experienced by their White British counterparts. This is consistent with the only previous study of this type in which predominantly written consultations were shown to people from different ethnic groups in the United States where the conclusion was also that differences in rating were more likely to represent differences in care than differences in expectation or scale use. [74]

There is a clear practice implication of this result: low scores from South Asian patients should be investigated as possible indicators of poor care. This is relevant to all settings, not just primary care.

3. Using data on patient experience for quality improvement

The results which we have discussed so far indicate that the results of patient experience surveys such as the GP Patient Survey can identify areas where there are important gaps in care that the NHS provides, such as patients being able to see a doctor of their choice. However, although patients tend to give very high scores for doctor-patient communication, these conceal significant negative experiences which patients describe and which independent observers can see in recorded primary care consultations. These issues extend to minority ethnic patients and our research suggests that the negative scores which South Asian patients record (compared to White British patients) do represent genuine problems with care. This therefore brings us to the important issue of how data from patient surveys can be used to improve care.

Current national approaches to measuring patient experience, including communication, rely on practice-level assessments of care. In chapter 9, we outline the results of a patient experience survey we conducted across 25 general practices, asking patients specifically about their experience of a particular consultation with a named GP. We found that practice-level scores for communication mask considerable variation between GPs within that practice, notably for those practices receiving poorer communication scores overall. Such “poorly performing” practices, which may be identified as such through the national GP Patient Survey, may in fact contain GPs with communication skills ranging from very poor to very good. This has important implications for the use of national survey data to identify primary care practices and practitioners in need of improvement.

In chapters 7 and 8 we describe the two studies in which we sought the views of GPs and practice staff on survey results, seeking to understand how they could better be used as quality improvement tools. Chapter 7 describes focus groups with practice staff following feedback of practice level scores on patient experience and chapter 8 describes interviews with GPs after we conducted a survey in which they got individual feedback from surveys returned by patients whom they had seen in surgeries. In chapter 11, we describe how out-of-hours providers use data from patient surveys.

Broadly, staff in different primary care settings neither believed nor trusted patient surveys. Concerns were expressed about the validity and reliability of surveys (some practices have very low rates of response) and of the likely representativeness of people who responded. Some practice groups mentioned recent negative experiences with pay linked to survey scores as part of the Quality and Outcomes Framework (a technicality of the payment schedule meant that payments could be reduced even though practice performance had improved). There was also a view expressed that some patients had unreasonable expectations: staff worked as hard as they could and couldn’t be expected to respond to all patients’ ‘wants’. Some practices did describe improvements that they had made as a result of survey results. Those which were easiest to engage with related to practices’ office functions such as appointment systems and telephone answering. Addressing an individual doctor's performance (for example, communication skills) was much more difficult. Out-of-hours service staff were also concerned that service users did not understand the complex care pathways within urgent care settings, and that this might lead to unrealistic expectations of what individual services were expected to deliver. Staff viewed surveys as necessary, but not sufficient. Clear preferences for more qualitative feedback to supplement survey scores were expressed as this provided more actionable data upon which to mount quality improvement initiatives.

The doctors we interviewed expressed markedly ambivalent views in discussing feedback from surveys. Whilst they had a number of concerns about individual doctor surveys (credibility, reliability, concerns about patient motivation), they also expressed positive views about the importance of patient feedback in monitoring and improving services.

These results led us to consider how patient feedback might be obtained in a way that would engage doctors more actively with patient survey results to stimulate quality improvement. We conducted a preliminary evaluation of real-time feedback (RTF), using touch screens that patients could use to leave feedback following a primary care consultation. RTF was selected as this technology to address some of the problems identified by our research, such as providing practice feedback on a much more regular basis (e.g. fortnightly), and allowing practices the opportunity to add questions of their own to the RTF survey to increase the relevance of the results to their service.

Since RTF has not been widely used, an exploratory RCT and qualitative study were conducted to answer questions about the feasibility of using RTF in real-world general practice, to estimate likely response rates, to get patient and staff views on providing feedback in this way, and to estimate the costs to a practice of introducing RTF. We also included facilitated feedback in one arm of the exploratory trial.

In our exploratory trial, only 2.5% of consulting patients left any RFT without prompting; however, if encouraged to use RTF by staff, as many as 60% of patients did so. Encouragement was rare, with such encouragement provided in only 5% of over 1100 patient-staff interactions that we observed in reception areas. Of patients who used RTF, 86% found it easy to use and were positive about it as a feedback method. Lack of awareness of the screens and lack of time were the commonest reasons for not giving feedback.

Staff were broadly positive about using RTF and practices valued the ability to include their own questions in the survey. Practices which had open communication between staff members tended to be more positive about using patient feedback. Practice staff identified clear benefits from having a facilitated session for discussion of patient feedback and having protected time to discuss the results.

Had practices not been taking part in a research study, the cost of RTF to practices would have been substantial at over £1000 for the 12 weeks, with the bulk relating to providing the equipment and analysing and feeding back data collected from the touch screens.

Although the absolute number of patients providing RTF feedback to each practice (>100) was comparable to the number of respondents per practices in the national GP Patient Survey, we do now know how the much lower response rate in our RTF study (2.5%) would have affected the actual results of the surveys (it was not part of our study design to find this out). We do not know how representative or valuable the views of a small proportion of patients responding are, just as we do not know how representative the views of the very small numbers of patients providing the narrative feedback that is recorded on NHS Choices.

Considering these results together, we have been able to identify some clear learning to take forward into a future clinical trial examining the potential utility and effectiveness of RTF in informing service delivery in primary care.

Implications for practice

The work that we have carried out over the past five years has clear implications for practice. We summarise these here.

1. The importance of patient experience

Our research supports the continuing emphasis on obtaining patient experience feedback as an important means of informing NHS care. Whilst continuing effort should be invested in refining the most effective and meaningful mechanism to capture high quality patient feedback, the key challenge is to provide primary care staff with the support and means to enable them to act upon patient feedback.

2. The need for action on the quality of care for minority ethnic groups

There has been much speculation as to whether the lower scores reported by minority ethnic groups on numerous patient experience surveys are “real”, reflecting poorer quality of care, or an artefact of the questionnaires used or higher expectations of care. We have now conducted a series of studies to progressively examine this issue to understand with greater certainty the major drivers of reported variations in care. Examinations of survey responses, interviews with patients, and an innovative experimental vignette study combine to strongly suggest that it is the former: patients from South Asian backgrounds experience considerably poorer communication with GPs than their White British counterparts. It is of concern that survey results may be dismissed as artefactual when, in fact, they are likely to point to real areas of concern. Effort should be invested to ensure lower scores from such groups on patient experience surveys in both primary care secondary care are investigated as markers of poorer quality of care.

3. Patients give over-positive responses when rating their care

Our results show the difficulty that patients have in feeding back negative experiences in questionnaire surveys. This suggests that there is more work to be done in improving patient experience than might be suggested by the high scores which are commonly seen in patient surveys. However, patients’ reluctance to criticise a doctor or provider with whom they have to maintain an ongoing relationship will not be addressed simply by changing the survey method. Efforts should be made to ensure providers and managers understand that absolute scores paint an optimistic picture of patients’ true views.

4. Surveys are not sufficient to fully capture patient feedback

Across primary and out-of-hours care settings, staff view patient surveys as necessary, but not sufficient. Alternative methods for gaining more qualitative feedback were commonly used to supplement survey scores, with free text often viewed as providing more actionable data than responses to standard survey questions. Taken alongside our findings on patients' reluctance to criticise doctors through surveys and staff challenges to the credibility of surveys, we suggest that additional approaches are therefore needed to better capture aspects of patient experience that can be used to improve quality of care.

5. The need for valid, reliable individual level feedback for doctors

Despite the comments above, we have shown that there is substantial variation in performance within practices for aspects of care related to individual doctors (e.g. doctor patient communication). Reporting patient experience at practice level masks this variation and makes it more difficult for doctors to relate to feedback. However, we have also shown that if a practice has overall high scores for doctor-patient communication, it is very unlikely that such a practice contains a low scoring doctor. In contrast, where a practice is low scoring, individual doctors may be high or low scoring. Therefore if there are additional requirements for individual level surveys, they could be focused on practices with low overall scores. Additionally, robust mechanisms are required to help practices, particularly lower scoring practices, identify and support individual doctors whose patient feedback identifies areas of potential improvement.

We note that, at present, data are provided at practice level for the GP Patient Survey, scores are produced at practice level for the Friends and Family Test, and GPs have to provide individual level surveys to meet GMC requirements for revalidation. These result in considerable overlap and duplication and add to the sense that these are 'boxes to be ticked' rather than sources of information that are valuable for improving care.

6. Patient surveys need to become more meaningful to staff

Our research shows that primary care staff in different settings are ambivalent about the value of patient surveys. While believing in general about the importance of issues such as doctor patient communication, they use every opportunity to challenge the credibility and reliability of scores produced by national surveys. This is not helped by their recent experiences, e.g. of a poorly conceived attempt to tie financial incentives to patient reports of waiting time to get an appointment, [281] and the imposition of the Friends and Family Test which is even regarded as limited value for comparing healthcare organisations by NHS England. [57]

On the whole, practices found it easier to engage with items on surveys that related to practice management (e.g. availability of appointments, ability to get through on the phone) than to issues around communication between patients and clinical staff. Staff viewed surveys as necessary, but not sufficient, and expressed a clear preference for qualitative feedback to supplement survey scores as this provided more actionable data upon which to mount quality improvement initiatives.

Immediacy of feedback, regularity of feedback, and having some control over the questions asked were all aspects of our experiment with real time feedback that were valued by practices and had the potential to make feedback more useful. However, a number of important questions remain before real time feedback could be recommended as a replacement for postal questionnaires. We outline these in the next section on research recommendations.

7. The value of surveys in monitoring national trends

Despite some reservations about the value of national surveys as vehicles for stimulating quality improvement in general practices and out-of-hours services, they can be important for monitoring national trends. For example, the GP Patient Survey is the only source of data which demonstrates that, year on year for the past five years, patients report than they have had increasing difficulty in seeing a doctor of their choice. Indeed, for out-of-hours services the GP Patient survey is the only way to monitoring such trends as individual services use very different tools and approaches precluding comparisons. Additionally, patient feedback – particularly in secondary care – is used for organisational risk assessment and regulatory monitoring. However, where national surveys are used to monitor trends in care it is important that the questions stay the same. In contrast to questions in the GP Patient Survey on patients being able to see a doctor of their choice, questions in the survey on access have undergone major changes making it difficult to follow long term trends. However, it should be noted that much smaller sample sizes are required to monitor national trends and comparable national surveys often number in tens of thousands of participants rather than millions. Our work on out-of-hours care suggests some ways in which the current questions in the GP Patient Survey could be improved.

8. Development of surveys in out-of-hours care

Our work on the use of patient experience surveys in out-of-hours care highlights a number of areas requiring consideration. National quality requirements (NQR 5) state all out-of-hours services must audit patient experience but provide no information on how to do this. [282] In the absence of clear guidance on tools and approaches, many services are taking different tacks to both collect and act upon patient feedback, with little comparison between services possible. As well as being inefficient in approach, with little consistency or shared learning, this also precludes national comparisons being made between providers. We suggest that NQR 5 should be reviewed and tightened to avoid the duplication of effort occurring in different services.

Secondly, out-of-hours items from the GP Patient Survey are now being used for the purposes of CQC and National Audit monitoring of out-of-hours care. Our research in this area commenced prior to the launch of the CQC, and providers knew little about the GP Patient Survey and expressed concern about the relevance of the out-of-hours items. Our research suggests that, subject to minor amendments, the GPPS is suitable for this kind of national monitoring of OOH care: indeed, it is the only current approach suitable for monitoring, given the variation in approaches to patient feedback currently taken by service providers. However, whilst the GP Patient Survey enables the use of benchmarking, it is not sufficiently detailed to support quality improvement and as such is unlikely to replace the in-house methods and tools being used by providers. We also note that current presentations of GP Patient Survey data for out-of-hours care are at ‘commissioner’ level: as providers often cover more than one commissioner level, such analyses may not highlight problems occurring at the larger organisational level. Finally, in order to look at the performance of different out-of-hours providers on key patient experience measures, it is important that NHS England maintains a list of such providers to ensure oversight, which it currently does not.

Overall, large scale postal surveys are likely to remain the dominant approach for gathering patient feedback for the time being, although refinements to this approach as well as the development of other modes are required to address the weaknesses we have identified. We are aware that providers are experimenting with a wide range of other approaches, one of which (RTF) has been part of our research. Other methods include interviews and focus groups, online feedback, analysis of complaints, practice participation groups and social media. In the following section, we outline recommendations for research, and identify the criteria that any new methods will need to meet in order to become useful quality improvement tools.

Recommendations for research

The world of patient feedback is becoming increasingly diverse and complex, with standard patient survey approaches being supplemented by the use of tablets, kiosks, online feedback including that provided by the NHS and by commercial organisations, analysis of complaints, the use of interviews and focus groups, and practice participation groups. In addition, social media may come to play an important part in how patients choose their doctor and how they feedback on their experiences. Some of these new approaches are being evaluated in terms of their ability to provide more detailed information on what is needed to improve services, for example by using patient narratives [283] and through the analysis of internet based feedback.[284, 285] However, despite the plethora of approaches to gathering patient feedback, our research demonstrates that there is a major deficit in taking action as a result of such feedback. Enabling and supporting providers to engage with and plan changes may require complex whole system approaches, and our knowledge of what is most effective in this area is currently sparse.

Research is therefore needed into how gathering and acting upon patient feedback may be best supported, across five key areas:

1. How patient experience can be captured so that it more effectively identifies areas of performance that could be improved. This should include investigation of diverse methods of obtaining patient feedback to support patients to highlight poor care where necessary. An additional important area of work is how some of the issues highlighted within this report – such as patients' reluctance to criticise – apply to different approaches to assessing patient experience using either rating-type or report-type questionnaire items.
2. The system, practitioner and patient factors which influence poorer reported experiences of care in South Asian patient groups, and how these may be addressed: this should include a particular focus on the impact of cross-cultural consultations.
3. How information from patients can be fed back to clinicians and services in a way that appears credible to them. This should include evaluations of approaches to increase the plausibility of patient surveys, such as greater use of bench-marking and innovative ways of presenting and interpreting findings, as well as assessment of

varying, tailored ways of presenting feedback to the different health care professionals who might receive feedback on their care. Of additional relevance here is how clinicians are encouraged to reflect on their own performance and others' assessments of this, with the aim of understanding where and how gaps in evaluations may occur.

4. How services can be organised and managed in such a way that patient feedback is seen as a positive opportunity for improving services.
5. What interventions are most effective in improving care where deficiencies in care are identified? The area where there is the greatest gap here is in doctor patient communication, where our results show that clinicians have great difficulty in even discussing deficiencies among their colleagues, and few effective interventions exist.

Our finding in the research on out-of-hours care that commercial providers had lower ratings on patient experience than services provided by the NHS is consistent with previous work suggesting that practices working under APMS contracts which are sometimes provided by the private sector may give worse care. [286] However, the circumstances in which commercial providers gain contracts for primary care services may be very different from those in other areas. The way in which the primary care workforce is configured is changing rapidly with an increase in the proportion of salaried GPs, the development of GP federations and super-practices and an increase in the number of large scale provider groups (owned both by commercial companies and by GPs). It is important that these changes should be monitored so that we understand their impact on quality of care.

Report word count

Total report word count, including Tables and Figures: 91,794

Acknowledgements

This report presents independent research funded by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research scheme (NIHR PGfAR - RP-PG-0608-10050). The views expressed in this publication are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

We would like to express our particular thanks to the lay members of the Improve Advisory Group for their commitment over the 5 years of the programme, and for sharing their experience, views and guidance throughout this time. Thanks too, to Dr Sahadev Swain for his contribution to the group and, in particular, his expertise on the issues facing patients from minority ethnic backgrounds.

Contribution of authors

The report authors' contributions are listed for each workstream in the tables below, alongside the contributions from the many others who made this programme of work possible. Chapter 1 (the Introduction) and Chapter 12 (the Conclusions) were drafted by John Campbell and Martin Roland, with additional input from Jenni Burt. The practice recommendations outlined by Martin Roland, John Campbell and Jenni Burt were discussed at a full Improve team meeting with particular contributions from Georgios Lyratzopoulos, Gary Abel, Suzanne Richards, Marc Elliott, Charlotte Paddison and Jenny Newbould. Julia Beckwith conducted a review of the literature which supported Chapter 1. Details of overall contribution from authors are listed in Table 56 below.

Table 56. Overall author contribution to programme of research

Overall Programme: Contributors			
Name and role		**Chapters contributed to**	**Nature of contribution**
Martin Roland	Professor of Health Services Research, University of Cambridge	All	Co-Chief Investigator and Principal Investigator. Oversaw the design and conduct of the study, the analysis and interpretation of the data, and edited the report as necessary
John Campbell	Professor of General Practice & Primary Care, University of Exeter	All	Co-Chief Investigator: contributed to the design and conduct of the study, the analysis and interpretation of the data, and edited the report as necessary.
Jenni Burt	Senior Research Associate, University of Cambridge	All	Programme Manager: contributed to the design and oversaw the conduct of the programme, reviewing the literature, analysis and interpretation of the data, and writing of individual project reports. Drafted and edited the final report as necessary
Jenny Newbould	Honorary Research Fellow, University of Cambridge	Chapters 2, 7, 8, 9, 10	Contributed to the design of the study, reviewing the literature, participant recruitment and data collection, analysis and interpretation of the data, and drafting the report. Contributed to qualitative data collection (10).
Nadia Llanwarne	Academic Clinical Fellow/ GP, University of Cambridge	Chapters 2, 6	Study researcher: responsible for reviewing the literature, participant recruitment and data collection, analysis and interpretation of the data, and drafting of the report.
Antoinette Davey	Research Fellow, University of Exeter	Chapters 2, 3, 7, 8, 9, 10	Study researcher: responsible for reviewing the literature, contributed to study organisation (Chapter 9), participant recruitment and data collection, analysis and interpretation of the data, and commented on the project report
Natasha Elmore	Research Assistant, University of Cambridge	Chapters 2, 3, 6, 7, 8, 9, 10	Study researcher: contributed to the design of the study, study set up (ethics and RMG), reviewing the literature, and assisted with participant recruitment and analysis of the data. Contributed to drafting of individual project reports and editing of final report.
Faraz Ahmed	PhD student, University of Cambridge	Chapter 2	Study researcher: contributed to study set up (ethics and RMG), participant recruitment, data collection, and analysis of the data for the South Asian workstream
Julia Beckwith	Research Assistant, University of	Chapters 2, 6	Study researcher: contributed to reviewing the literature and analysis of the data. Contributed to drafting the final report.

	Cambridge		
Hena Wali Haque	Research Assistant, University of Cambridge	Chapter 2	Study researcher: contributed to participant recruitment and data collection
Inocencio Maramba	Associate Research Fellow, University of Exeter	Chapters 2, 3, 7, 8, 9	Study researcher: assisted with study set up, practice and participant recruitment, responsible for local data entry systems and storage (3 only)
Emily Taylor	Research Associate, University of Cambridge	Chapters 2, 3, 7, 8, 9	Study researcher: assisted with study set up, participant recruitment and data collection
Gary Abel	Senior Research Associate, University of Cambridge	Chapters 2, 3, 4, 5, 6, 9	Lead statistician: contributed to the design of the study, conceived and conducted data analysis and interpretation, and drafted and edited the report as necessary.
Ahmed Aboulghate	PhD student, University of Cambridge	Chapter 4	Statistical analysis and interpretation of the data
Anthea Asprey	Associate Research Fellow, University of Exeter	Chapters 7, 9	Study researcher: conducted focus groups, analysed the data.
Heather Barry	Associate Research Fellow, University of Exeter	Chapter 11	Study researcher: contributed to study design, organisation, data collection, and analysis and interpretation of the findings, and drafting the project report.
John Benson	Senior Lecturer, University of Cambridge	Chapters 3, 6	Study advisor: contributed to the design of the study, the development of the GCRS and data interpretation
Olga Boiko	Associate Research Fellow, University of Exeter	Chapters 7, 8	Study researcher: contributed to the design of the study, study set-up, conducted focus groups, analysed the data and edited the project report as necessary (7 only).
Pete Bower	Professor, University of Manchester	Chapter 10	Study advisor: contributed to study design and reviewing the literature
Raff Calitri	Research Fellow, University of Exeter	Chapter 11	Study researcher: assisted with organisation, data collection, data analysis and interpretation
Mary Carter	Associate Research Fellow, University of Exeter	Chapters 9, 10	Study researcher: Collected qualitative and quantitative data; conducted the the majority of the qualitative analysis and contributed to the interpretation of the results; drafted the qualitative section of the project report
Marc N. Elliott	Distinguished Chair in Statistics, RAND Corporation	Chapters 3, 4, 5, 6, 9	Statistical advisor: contributed to the design of the study, contributed to data analysis and interpretation, and contributed to the drafting of

			the reports
Conor Farrington	Research Associate, University of Cambridge	Chapters 7, 8	Study researcher: contributed to the design of the study, study set up, data collection, analysed and interpreted the data, wrote the project report (8 only).
William Henley	Professor of Medical Statistics	Chapter 11	Study contributor: assisted with data analysis and interpretation
Val Lattimer	Professor of Health Services Research and Dean of Health Sciences, UEA	Chapter 4	Study advisor: formulated aspects of the research question, contributed to the design of the analysis, contributed to the interpretation of the data, and commented on the report
Cathy Lloyd	Professor of Health Studies, The Open University	Chapters 5, 6	Study advisor: contributed to the interpretation of the ethnicity interaction analysis
Georgios Lyratzopoulos	Reader in Cancer Epidemiology, UCL	Chapter 5	Lead for the first ethnicity analysis: designed and oversaw the analysis, interpreted the data, and contributed to the drafting of the report
Luke Mounce	Associate Research Fellow, University of Exeter	Chapter 10	Study statistician: Conducted quantitative analysis and contributed to the drafting of the project report
Charlotte Paddison	Senior Lecturer, Anglia Ruskin University	Chapters 3, 9	Study advisor: contributed to the design of the study, study set up (9 only) and data interpretation
Richard Parker	Research Assistant, University of Cambridge	Chapter 4	Statistical analysis and interpretation of the data
Suzanne Richards	Senior Lecturer in Primary Care	Chapter 11	Workstream lead: responsible for protocol design, study organisation, data collection, analysis and interpretation of the findings, drafting the report and critiquing all outputs for important intellectual content.
Martin Roberts	Senior Psychometrician, University of Plymouth	Chapter 9	Statistician: conducted data analysis and contributed to the drafting of the project report.
Claude Setodji	Senior Statistician, RAND Corporation	Chapter 5	Study statistician: designed and conducted the DIF analysis, interpreted the data, and contributed to the drafting of the report
Jonathan Silverman	Associate Dean, School of Clinical Medicine, University of Cambridge	Chapter 3	Study advisor: contributed to the design of the study, the development of the GCRS and data interpretation
Fiona Warren	Lecturer in Medical Statistics, University of Exeter	Chapter 11	Study contributor: assisted with data analysis and interpretation
Ed Wilson	Senior Research Associate,	Chapter 10	Study researcher: responsible for the economic

	University of Cambridge		analysis
Christine Wright	Research Fellow, University of Exeter Medical School	Chapter 10	Workstream lead. Contributed to study design (development of protocol), study organisation and set up (recruitment of practices, ethics and RMG), reviewing the literature, supervision of data collection, collation of cost analysis data, contributed to draft quantitative analysis plan, interpretation, drafting the project report.

There are a number of papers arising from the programme of research which are in preparation, in submission or under review. Those that have been published to date are listed in the acknowledgements sections of the appropriate Chapters, alongside the conference presentations undertaken.

Data archiving and sharing

Data arising from the various strands of this programme of work has been archived in accordance with the agreements covered by ethical approvals for the research. For Chapters 2 and 3, consent to participate was given on condition of video-recording of consultations and associated data being accessible only to members of the immediate Improve research team, therefore the data generated in these chapters is not suitable for sharing. GP Patient Survey data used in Chapters 4 and 5 are accessible via request to NHS England, please contact the author for further information. Chapter 6 data – rating of simulated consultations – are available via request to the author, as are survey data set out in Chapter 9. Data used in Chapters 7, 8, 10, and 11, including interviews, observations and focus groups, are not currently available to ensure confidentiality of practices who participated.

Acknowledgements

Chapter 2 - How do patients' respond to communication items on patient experience questionnaires? Video elicitation interviews with patients.

We would like to thank the thirteen general practices who participated in the project, particularly the 45 GPs who so kindly agreed for their consultations with patients to be video recorded. We would particularly like to thank the 52 patients who so generously gave up their time to take part in a video elicitation interview, without which this study would not have been possible. We also give our thanks to the three practices and eight GPs who took part in the additional interviews with South Asian respondents, and the 23 patients from these three practices who so generously gave up their time to take part in a video elicitation interview, without which this study would not have been possible.

Finally, we would like to thank the Improve Advisory Group who gave their advice on study design and interpretation of the data.

Table 57. Chapter 2 acknowledgements. Contributions to the video elicitation interview workstream

Chapter 2: Contributors

Name and role		Nature of contribution
Martin Roland	Professor of Health Services Research, University of Cambridge	Co-Chief Investigator and Principal Investigator for this work stream. Oversaw the design and conduct of the study, the analysis and interpretation of the data, and edited the report as necessary
John Campbell	Professor of General Practice & Primary Care, University of Exeter	Co-Chief Investigator: contributed to the design and conduct of the study, the analysis and interpretation of the data, and edited the report as necessary
Jenni Burt	Senior Research Associate, University of Cambridge	Programme Manager: contributed to the design and oversaw the conduct of the study, reviewing the literature, analysis and interpretation of the data. Drafted and edited the report as necessary
Jenny Newbould	Honorary Research Fellow, University of Cambridge	Workstream lead: contributed to the design of the study, reviewing the literature, participant recruitment and data collection, analysis and interpretation of the data, and drafting the report
Nadia Llanwarne	Academic Clinical Fellow/ GP, University of Cambridge	Study researcher: responsible for reviewing the literature, participant recruitment and data collection, analysis and interpretation of the data, and drafting of the report.

Antoinette Davey	Research Fellow, University of Exeter	Study researcher: responsible for reviewing the literature, participant recruitment and data collection, analysis and interpretation of the data, and commented on the report
Natasha Elmore	Research Assistant, University of Cambridge	Study researcher: contributed to the design of the study, study set up (ethics and RMG), reviewing the literature, and assisted with participant recruitment and analysis of the data
Faraz Ahmed	PhD student, University of Cambridge	Study researcher: contributed to study set up (ethics and RMG), participant recruitment, data collection, and analysis of the data for the South Asian workstream
Julia Beckwith	Research Assistant, University of Cambridge	Study researcher: contributed to reviewing the literature and analysis of the data
Hena Wali Haque	Research Assistant, University of Cambridge	Study researcher: contributed to participant recruitment and data collection
Inocencio Maramba	Associate Research Fellow, University of Exeter	Study researcher: assisted with participant recruitment
Emily Taylor	Research Associate, University of Cambridge	Study researcher: assisted with participant recruitment and data collection

Table 58. Chapter 2 acknowledgements. Conference presentations relating to the video elicitation interview workstream

Chapter 2: Conference presentations

Presenter	Title	Location and Date
Jenny Newbould	Patient questionnaires – a useful reflection of patient experience?	Society of Academic Primary Care London & South East Regional Conference, Cambridge, January 2015
Jenny Newbould	The tick and the talk: do patients' survey responses relate to their narrated experience of primary care consultations?	Society of Academic Primary Care Annual Scientific Meeting, Oxford, July 2015

Acknowledgements

Chapter 3 - The association between patients', raters' and GPs' assessment of communication in a consultation

We would like to thank the thirteen general practices who participated in the project, particularly the 45 GPs who so kindly agreed for their consultations with patients to be video recorded, and the receptionists and administrative staff who assisted with patient identification and recruitment. We would also like to thank the 529 patients who completed a study questionnaire following their consultation with a GP. The rating of the consultations would not have been possible without our trained GCRS raters, to whom we extend our thanks for their expertise. Pete Bower gave important advice on study design. Finally, we would like to thank the Improve Advisory Group who gave their advice on study design and interpretation of the data.

Table 59. Chapter 3 acknowledgements. Contributions to the consultation rating workstream

Chapter 3: Contributors

Name and role		Nature of contribution
Martin Roland	Professor of Health Services Research University of Cambridge	Co-Chief Investigator and Principal Investigator for this work stream. Oversaw the design and conduct of the study, the analysis and interpretation of the data, and edited the report as necessary
John Campbell	Professor of General Practice & Primary Care University of Exeter	Co-Chief Investigator: contributed to the design and conduct of the study, the analysis and interpretation of the data, and edited the report as necessary
Jenni Burt	Senior Research Associate University of Cambridge	Programme Manager and workstream lead: contributed to the design of the study, oversaw the conduct of the study, assisted with the analysis, interpreted the data, and drafted the report
Gary Abel	Senior Research Associate University of Cambridge	Lead study statistician: contributed to the design of the study, conducted data analysis, interpreted the data, and contributed to the drafting of the report
Marc Elliott	Distinguished Chair in Statistics RAND Corporation	Statistical advisor: contributed to the design of the study, contributed to data analysis and interpretation, and contributed to the drafting of the report
Natasha Elmore	Research Assistant University of Cambridge	Study researcher: contributed to the design of the study, study set up (ethics and RMG), assisted with participant recruitment, coordinated the rating of videos, responsible for local data entry systems and storage
Jenny Newbould	Honorary Research Fellow University of Cambridge	Study researcher: responsible for participant recruitment and data collection, assisted with data entry, storage and management, contributed to data interpretation
Nadia Llanwarne	Academic Clinical Fellow, University of Cambridge	Study researcher: responsible for participant recruitment and data collection, assisted with data entry, storage and management, contributed to data interpretation
Antoinette Davey	Research Fellow	Study researcher: responsible for participant recruitment and data

	University of Exeter	collection, assisted with data entry, storage and management
James Brimicombe	Data Manager University of Cambridge	Data manager: designed data entry systems and storage
Inocencio Maramba	Associate Research Fellow University of Exeter	Study researcher: assisted with participant recruitment, responsible for local data entry systems and storage
Emily Taylor	Research Associate University of Cambridge	Study researcher: assisted with participant recruitment and data collection
Charlotte Paddison	Senior Lecturer, Anglia Ruskin University	Study advisor: contributed to the design of the study and data interpretation
John Benson	Senior Lecturer University of Cambridge	Study advisor: contributed to the design of the study, the development of the GCRS and data interpretation
Jonathan Silverman	Associate Dean, School of Clinical Medicine, University of Cambridge	Study advisor: contributed to the design of the study, the development of the GCRS and data interpretation

Table 60. Chapter 3 acknowledgements. Conference presentations arising from the consultation rating workstream

Chapter 3: Conference presentations

Presenter	Title	Location and Date
Gary Abel	Rating communication in GP consultations: do patients and trained experienced raters agree?	Society for Academic Primary Care London & South East Regional Conference, Cambridge, January 2015
Gary Abel	Rating communication in GP consultations: do patients and experienced trained raters agree?	Health Services Research Network (HSRN), July 2015
Gary Abel	Rating communication in GP consultations: do patients and experienced trained raters agree?	Society of Academic Primary Care Annual Scientific Meeting, Oxford, July 2015

Acknowledgements
Chapter 4 – Seeing a clinician of the patient's choice

This chapter represents a series of analyses of GP Patient Survey data conducted by members of the Improve team during the course of the programme. We would like to thank Ipsos MORI for assisting with various enquiries during this time, as well as NHS England for their continued support for our analytical work in this area.

Table 61. Chapter 4 acknowledgements. Contributions to the GP Patient Survey analyses

Chapter 4: Contributors		
Name and role		**Nature of contribution**
Martin Roland	Professor of Health Services Research, University of Cambridge	Co-Chief Investigator and Principal Investigator for this work stream. Oversaw the design and conduct of the analyses, the interpretation of the data, and edited the report as necessary
John Campbell	Professor of General Practice and Primary Care University of Exeter Medical School	Co-Chief Investigator: contributed to the design and conduct of the analysis and interpretation of the data, and edited the report as necessary
Gary Abel	Senior Research Associate University of Cambridge	Lead study statistician: conceived and conducted data analyses, interpreted the data, and drafted the report
Ahmed Aboulghate	PhD student	Statistical analysis and interpretation of data
Jenni Burt	Senior Research Associate University of Cambridge	Programme Manager: contributed to the design and conduct of the analysis, the interpretation of the data, and edited the report as necessary
Marc Elliott	Senior statistician, RAND US	Statistical analysis and interpretation of data
Richard Parker	Research Assistant, University of Cambridge	Statistical analysis and interpretation of data
Val Lattimer	Professor of Health Services Research and Dean of Health Sciences, UEA	Study advisor: formulated aspects of the research question, contributed to the design of the analysis, contributed to the interpretation of the data, and commented on the report

Table 62. Chapter 4 acknowledgements. Published papers arising from the GP Patient Survey analyses

Chapter 4: Published papers		
Authors	**Title**	**Journal**
Aboulghate, A, Abel, G, Elliott, MN, Parker, RA, Campbell, J, Lyratzopoulos, G, Roland, M	Do English patients want continuity of care, and do they receive it?	*BJGP* 2012; 62: e567-75

Acknowledgements

Chapter 5 - Analyses of GP Patient Survey data to explore variations in patient experience by ethnic group and practice

This chapter represents further analyses of GP Patient Survey data conducted by members of the Improve team during the course of the programme. We would like to thank Ipsos MORI for assisting with various enquiries during this time, as well as NHS England for their continued support for our analytical work in this area.

Table 63. Chapter 5 acknowledgements. Contributions to the GP Patient Survey analyses on minority ethnic experience

Chapter 5: Contributors

Name and role		**Nature of contribution**
Martin Roland	Professor of Health Services Research University of Cambridge	Co-Chief Investigator and co-principal investigator for this work stream: oversaw the design and conduct of the data analysis and interpretation of the data, and edited the report as necessary
John Campbell	Professor of General Practice & Primary Care University of Exeter	Co-Chief Investigator and co-principal investigator for this workstream: oversaw the design and conduct of the data analysis and interpretation of the data, and edited the report as necessary
Jenni Burt	Senior Research Associate University of Cambridge	Programme Manager: contributed to the design of a number of the ethnicity analyses, contributed to the interpretation of the data, and drafted the report
Gary Abel	Senior Research Associate University of Cambridge	Study statistician: designed and conducted the ethnicity interaction analysis and the practice-level analysis interpreted the data, and drafted the report
Marc Elliott	Distinguished Chair in Statistics RAND Corporation	Statistical advisor: contributed to the design, conduct and interpretation of the analyses, and contributed to the drafting of the report
Georgios Lyratzopoulos	Reader in Cancer Epidemiology UCL	Lead for the first ethnicity analysis: designed and oversaw the analysis, interpreted the data, and contributed to the drafting of the report
Claude Setodji	Senior Statistician RAND Corporation	Study statistician: designed and conducted the DIF analysis, interpreted the data, and contributed to the drafting of the report
Cathy Lloyd	Professor of Health Studies The Open University	Study advisor: contributed to the interpretation of the ethnicity interaction analysis

Table 64. Chapter 5 acknowledgements. Conference presentations arising from GPPS minority ethnic groups analyses

Chapter 5: Conference presentations		
Presenter	**Title**	**Location and Date**
Jenni Burt	Do ethnic disparities in patient reported GP-patient communication vary by age and gender? Evidence from a national patient survey	Society of Academic Primary Care Annual Scientific Meeting, Oxford, July 2015

Table 65. Chapter 5 acknowledgements. Published papers arising from GPPS minority ethnic groups analyses

Chapter 5: Published papers		
Authors	**Title**	**Journal**
Lyratzopoulos G, Elliott M, Barbiere JM, Henderson A, Staetsky L, Paddison C et al.	Understanding ethnic and other socio-demographic differences in patient experience of primary care: evidence from the English General Practice Patient Survey	*BMJ Quality & Safety.* 2012;21(1):21-9
Setodji CM, Elliott MN, Abel G, Burt J, Roland M, Campbell J	Differential Item Functioning in the English General Practice Patient Survey: Comparison of South Asian and White British Subgroups	*Medical Care* 2015. 53(9): 809-817
Burt J, Lloyd C, Campbell J, Roland, M and Abel, G.	Variations in GP-patient communication by ethnicity, age, and gender: evidence from a national primary care patient survey	*British Journal of General Practice 2016. 66(642):e47-52*

Acknowledgements

Chapter 6 - How do White British and Pakistani people rate communication within simulated GP-patient consultations? Experimental vignette study

This complex study would not have been possible without the contributions of many people. We would particularly like to thank Steve Attmore for his assistance with the development and recording of the study vignettes, and the Media Studio at Cambridge University Hospitals NHS Trust for their assistance with the recording and editing of the vignettes. We would also like to thank all the actors who took roles in the vignettes. A wide range of staff at Ipsos MORI contributed to developing and refining the study design: we would particularly like to thank Anna Carluccio and Lara Sarson for their professional oversight of the project, and their colleagues James Wilks and Victoria Hough for their input. Pete Bower gave important advice on study design. Special thanks go to the team of Ipsos MORI fieldworkers who so diligently worked to recruit participants and conduct interviews. Special thanks, too, to the 1,124 participants who gave up their time to view and rate the simulated GP consultation vignettes – without them this project would not have been possible. Finally, many thanks to the Improve Advisory group for their assistance with study design and data interpretation.

Table 66. Chapter 6 acknowledgements. Contributions to the vignette study

Chapter 6: Contributors		
Name and role		**Nature of contribution**
Martin Roland	Professor of Health Services Research University of Cambridge	Co-Chief Investigator and Principal Investigator for this work stream. Oversaw the design and conduct of the study, the analysis and interpretation of the data, and edited the report as necessary
John Campbell	Professor of General Practice & Primary Care University of Exeter	Co-Chief Investigator: contributed to the design and conduct of the study, the analysis and interpretation of the data, and edited the report as necessary
Jenni Burt	Senior Research Associate University of Cambridge	Programme Manager & workstream lead: designed the study, designed & produced the study vignettes, oversaw study conduct, assisted with the analysis & interpretation of the data, & drafted the report
Gary Abel	Senior Research Associate University of Cambridge	Lead study statistician: designed the study, conducted data analysis, interpreted the data, and drafted the report
Marc Elliott	Distinguished Chair in	Statistical advisor: contributed to the design of the study, contributed to data analysis and interpretation, and contributed to the drafting of the

	Statistics RAND Corporation	report
Natasha Elmore	Research Assistant University of Cambridge	Study researcher: conducted review of the literature, contributed to the development and production of study vignettes
Cathy Lloyd	Professor of Health Studies The Open University	Study advisor: contributed to the study design and interpretation of the data
John Benson	Senior Lecturer University of Cambridge	Study advisor: contributed to the development and production of the study vignettes
Nadia Llanwarne	Academic Clinical Fellow, University of Cambridge	Study advisory: contributed to the production of the study vignettes
Julia Beckwith	Research Assistant University of Cambridge	Study researcher: coordinated the rating of vignettes by GCRS raters
Lara Sarson	Research Manager Ipsos MORI	Project manager at market research agency: contributed to study design and managed data collection
Anna Carluccio	Research Director Ipsos MORI	Project leader at market research agency: contributed to study design and oversaw data collection
Steve Attmore	Simulated patient co-ordinator University of Cambridge	Study contributor: assisted with the design and development of the vignettes; recruited all vignette actors; oversaw vignette roleplays

Table 67. Chapter 6 acknowledgements. Conference presentations arising from the vignette study

Chapter 6: Conference presentations		
Presenter	**Title**	**Location and Date**
Jenni Burt	How do White British and Pakistani people rate communication within simulated GP-patient consultations? A national experimental vignette study	Society of Academic Primary Care Annual Scientific Meeting, Oxford, July 2015
Jenni Burt	How do White British and Pakistani people rate communication within simulated GP-patient consultations? A national experimental vignette study	Health Services Research Network (HSRN), July 2015
Jenni Burt	Using Vignettes to Understand Differences in Patient Experiences for White and Pakistani Adults in England	North American Primary Care Research Group Annual Meeting, Cancun, October 2015

Acknowledgements

Chapter 7 - Attitudes to receiving feedback from patient experience surveys: focus groups with practice staff

This study, involving focus groups with practice staff, was linked to the survey reported in Chapter 9. We would like to thank all those practices who made our research team so welcome throughout the data collection period. We would particularly like to thank all the practice staff who gave up their time to take part in focus groups, and reflect on the way in which they engage with patient feedback: we are most grateful for their important contribution to this work. Thanks, too, to the Improve Advisory group who supported the research team throughout.

Table 68. Chapter 7 acknowledgements. Contributions to the focus group study

Chapter 7: Contributors		
Name and role		**Nature of contribution**
John Campbell	Professor of General Practice & Primary Care University of Exeter	Co-Chief Investigator and Principal Investigator for this work stream. Oversaw the design and conduct of the study, the analysis and interpretation of the data, and drafted the report
Martin Roland	Professor of Health Services Research University of Cambridge	Co-Chief Investigator: contributed to the design and conduct of the study, the analysis and interpretation of the data, and edited the report as necessary
Jenni Burt	Senior Research Associate University of Cambridge	Programme Manager: contributed to the design of the study, oversaw the conduct of the study, analysed the data, and drafted the report
Olga Boiko	Associate Research Fellow University of Exeter	Study researcher: contributed to the design of the study, study set-up, conducted focus groups, analysed the data, and edited the report as necessary
Natasha Elmore	Research Assistant University of Cambridge	Study researcher: assisted with study set-up, practice recruitment, data collection, and data entry
Antoinette Davey	Research Fellow University of Exeter	Study researcher: assisted with study set-up, practice recruitment, and data entry
Jenny Newbould	Honorary Research Fellow University of Cambridge	Study researcher: assisted with study set-up, practice recruitment, and data collection
Conor Farrington	Research Associate	Study researcher: assisted with data collection

	University of Cambridge	
Emily Taylor	Research Associate University of Cambridge	Study researcher: assisted with study set-up, practice recruitment, and data collection
Inocencio Maramba	Associate Research Fellow University of Exeter	Study researcher: assisted with study set-up and practice recruitment
Anthea Asprey	Associate Research Fellow University of Exeter	Study researcher: conducted focus groups, analysed the data.

Table 69. Chapter 7 acknowledgements. Conference presentations arising from the focus group study

Chapter 7: Conference presentations		
Presenter		**Location and Date**
Olga Boiko	Acting on patient feedback: managing change and improving service in response to patient surveys in primary care.	Society for Academic Primary Care South West meeting, March 2013

Table 70. Chapter 7 acknowledgements. Published papers arising from the focus group study

Chapter 7: Published papers		
Authors	**Title**	**Journal**
Boiko O, Campbell JL, Elmore N, Davey AF, Roland M, Burt J.	The role of patient experience surveys in quality assurance and improvement: a focus group study in English general practice.	*Health Expectations*. 2014.

Acknowledgements

Chapter 8 - Attitudes to receiving feedback from patient experience surveys: interviews with GPs

This study, involving interviews with GPs, was linked to the survey reported in Chapter 9. We would like to thank all those practices who made our research team so welcome as we conducted the survey. The study would not have happened without those GPs who were willing to participate in interviews with our research team, in which they reflected honestly and openly about their views on patient feedback and, in particular, their personal experience of receiving patient feedback. Thanks go additionally to the Improve Advisory group who supported the research team throughout.

Table 71. Chapter 8 acknowledgements. Contributions to the GP interview study

Chapter 8: Contributors		
Name and role		**Nature of contribution**
John Campbell	Professor of General Practice & Primary Care University of Exeter	Co-Chief Investigator and Principal Investigator for this work stream. Oversaw the design and conduct of the study, the analysis and interpretation of the data, and edited the report as necessary
Martin Roland	Professor of Health Services Research University of Cambridge	Co-Chief Investigator and Principal Investigator for this work stream: oversaw the design and conduct of the study, the analysis and interpretation of the data, and edited the report as necessary
Conor Farrington	Research Associate University of Cambridge	Study researcher: contributed to the design of the study, study set-up, carried out data collection, analysed and interpreted the data, and wrote the report
Jenni Burt	Senior Research Associate University of Cambridge	Programme Manager: contributed to the design of the study, oversaw the conduct of the study, contributed to data collection and the analysis of the data, and edited the report as necessary
Olga Boiko	Associate Research Fellow	Study researcher: contributed to the design of the study, study set-up, carried out data collection, and contributed to the analysis of the data
Natasha Elmore	Research Assistant University of Cambridge	Study researcher: assisted with study set-up and practice recruitment
Antoinette Davey	Research Fellow University of Exeter	Study researcher: assisted with study set-up and practice recruitment, and data entry
Jenny Newbould	Honorary Research	Study researcher: assisted with study set-up and practice recruitment

	Fellow University of Cambridge	
Emily Taylor	Research Associate University of Cambridge	Study researcher: assisted with study set-up and practice recruitment
Inocencio Maramba	Associate Research Fellow University of Exeter	Study researcher: assisted with study set-up and practice recruitment

Table 72. Chapter 8 acknowledgements. Conference presentations arising from the GP interview study

Chapter 7: Conference presentations		
Presenter		**Location and Date**
Conor Farrington	Dimensions of Ambivalence: Doctors and Patient Experience Surveys in Primary and Secondary Care	British Sociological Association Medical Sociology Group Annual Conference, University of York, September 2015

Acknowledgements

Chapter 9 - Understanding high and low patient experience scores: analysis of patients’ survey data for general practices and individual GPs

The survey reported within this Chapter involved a large scale data collection effort to which many people contributed. We would like to extend particular thanks to the 25 practices who participated in the study, and the staff within these who went out of their way to ensure the survey was conducted efficiently and to the highest standards. Thanks, too, to all those patients who took the time to respond to the questionnaire they received. The Improve Advisory group contributed particular help with devising study documentation, for which we are very grateful. Finally, Pete Bower gave important advice on study design.

Table 73. Chapter 9 acknowledgements. Contributions to the patient survey workstream

Chapter 9: Contributors		
Name and role		**Nature of contribution**
John Campbell	Professor of General Practice & Primary Care University of Exeter	Co-Chief Investigator and Principal Investigator for this work stream. Oversaw the design and conduct of the study, the analysis and interpretation of the data, and edited the report as necessary
Martin Roland	Professor of Health Services Research University of Cambridge	Co-Chief Investigator: oversaw the design and conduct of the study, the analysis and interpretation of the data, and edited the report as necessary
Jenni Burt	Senior Research Associate University of Cambridge	Programme Manager: contributed to the design of the study, oversaw study set up, practice recruitment and the conduct of the study, contributed to data collection, analysis and interpretation, and drafted and edited the report as necessary
Gary Abel	Senior Research Associate University of Cambridge	Statistician: contributed to the study design, sampled practices for the study, contributed to the analysis plan, contributed to the analysis and data visualisation and interpretation of the study, and helped to draft and edit the report
Antoinette Davey	Research Fellow University of Exeter	Study researcher: contributed to the organisation of the study, practice recruitment, data collection, and drafting of the report
Martin Roberts	Senior Psychometrician University of Plymouth	Statistician: conducted data analysis and contributed to the drafting of the report
Natasha Elmore	Research Assistant University of Cambridge	Study researcher: contributed to study set-up, practice recruitment, data collection and data entry
Inocencio Maramba	Associate Research Fellow University of Exeter	Study researcher: contributed to data collection
Mary Carter	Associate Research Fellow University of Exeter	Study researcher: up contributed to study set-up, practice recruitment, and data collection
Marc Elliott	Distinguished Chair in Statistics RAND Corp	Statistical advisor: contributed to the study design, analysis plan, analysis and data visualisation and interpretation of the study, and helped to draft and critique the report
Jenny Newbould	Honorary Research Fellow University of Cambridge	Study researcher: contributed to study set-up, practice recruitment, and data collection
Emily Taylor	Research Associate	Study researcher: contributed to study set-up, practice recruitment, and data collection

	University of Cambridge	
Anthea Asprey	Associate Research Fellow University of Exeter	Study researcher: contributed to study set-up
Emma Whitton	Administrator University of Exeter	Study administrator: assisted with study organisation and data entry
Amy Gratton	Administrator University of Exeter	Study administrator: assisted with study organisation and data entry
Dawn Swancutt	Project Manager University of Exeter	Project manager: assisted with study set-up
Charlotte Paddison	Senior Research Associate University of Cambridge	Study advisor: assisted with study design and set up

Table 74. Chapter 9 acknowledgements. Conference presentations arising from the patient survey workstream

Chapter 9: Conference presentations

Presenter	Title	Location and Date
Antoinette Davey	Investigating the stability of patient responses to GP patient survey items using test retest methodology.	Society of Academic Primary Care Annual Scientific Meeting, Oxford, July 2013
Jenni Burt	My GP practice scores well on doctors' consultations skills – so why doesn't my doctor listen to me?	Society of Academic Primary Care Annual Scientific Meeting, Oxford, July 2013
Inocencio Maramba	Researching with our head in the clouds? Using tag clouds to analyse free text patient feedback	Medicine 2.0, London, September 2013

Table 75. Chapter 9 acknowledgements. Publications arising from the patient survey workstream

Chapter 9: Published papers		
Authors	**Title**	**Journal**
Roberts MJ, Campbell JL, Abel GA, Davey AF, Elmore NL, Maramba I, et al..	Understanding high and low patient experience scores in primary care: analysis of patients' survey data for general practices and individual doctors	*BMJ.* 2014;349.
Maramba ID, Davey A, Elliott M, Roberts M, Roland M, Brown F, Burt J, Boiko O, Campbell J.	Web-based Textual Analysis of Free-Text Comments from Patients in Primary Care	*JMIR Medical Informatics* 2015 3(2):e20.

Acknowledgements

Chapter 10 - Exploratory trial of a real-time feedback intervention to improve patient experience in general practice

The completion of this complex workstream was only possible with the contribution of a great many people. We would particularly like to thank the patients and staff from the twelve general practices who participated in the project. We thank also the facilitators who delivered team feedback reflection sessions at six practices during the feasibility and exploratory trial phases. We are grateful to staff from Customer Research Technology (CRT) Limited (particularly Richard Farrell, Toby Knight and Nicky Allen) who provided the touch-screen equipment, organised data cleaning and summarising, prepared RTF reports for practices, and provided technical assistance and advice before and during the RTF implementation period. Antoinette Davey and Mary Carter conducted all fieldwork in the South West of England. Natasha Elmore, Jenny Newbould, and Jenni Burt conducted fieldwork at the two Cambridge practices during the exploratory trial phase. Ed Wilson designed, analysed and reported the cost analysis elements. Luke Mounce conducted and reported the quantitative analysis for the exploratory trial phase. John Campbell, Martin Roland, Jenni Burt and Gary Abel provided helpful comments during the development of the study protocol. Chris Wright managed the overall conduct and delivery of the workstream.

The development of the Value of Patient Feedback tool as part of this workstream rested on the contribution of a wide range of people. We would particularly like to thank Nadia Llanwarne, John Benson, Felix Greaves and Pete Bower for their critical input into the construction of the scale; the GPs who took part in cognitive interviews; and all the clinicians who participated in the piloting phases of development by completing the scale in its various developmental incarnations. James Brimicombe was instrumental in coordinating online piloting of the tool, and we extend our particular thanks to him for all his support with this.

Table 76. Chapter 10 acknowledgements. Contributions to the RTF exploratory trial workstream

Chapter 10: Contributors		
Name and role		**Nature of contribution**
John Campbell	Professor of General Practice & Primary Care University of Exeter	Co-Chief Investigator and Principal Investigator for this work stream. Oversaw the design and conduct of the study, the analysis and interpretation of the data, and edited the report as necessary
Martin Roland	Professor of Health Services Research University of Cambridge	Co-Chief Investigator: oversaw the design and conduct of the study, the analysis and interpretation of the data, and edited the report as necessary
Christine Wright	Research Fellow University of Exeter Medical School	Workstream lead. Contributed to study design (development of protocol), study organisation and set up (recruitment of practices, ethics and RMG), reviewing the literature, supervision of data collection, collation of cost analysis data, contributed to draft quantitative analysis plan, interpretation, drafting the report.
Mary Carter	Associate Research Fellow University of Exeter	Study researcher. Collected quantitative and qualitative data; conducted the majority of the qualitative analysis and contributed to the interpretation of the results; drafted the qualitative section of the report.
Luke Mounce	Associate Research Fellow University of Exeter	Study statistician. Conducted quantitative analysis and contributed to the drafting of the report.
Antoinette Davey	Research Fellow University of Exeter	Contributed to the study design, organisation of the study, recruitment of practices, data collection & input, analysis & interpretation of qualitative data, and drafting the qualitative results for the report
Natasha Elmore	Research Assistant University of Cambridge	Study researcher. Responsible for local study set up (RMG and recruitment). Collected the data for Cambridge sites; entered the data for Cambridge sites; commented on the drafting of the report.
Jenni Burt	Senior Research Associate University of Cambridge	Programme manager: contributed to the design of the study, the qualitative analysis plan, conducted qualitative data collection, assisted with interpretation of the data; and edited the report as necessary.
Jenny Newbould	Honorary Research Fellow, University of Cambridge	Study researcher: contributed to qualitative data collection
Ed Wilson	Senior Research Associate University of Cambridge	Study researcher: responsible for the economic analysis
Pete Bower	Professor, University of Manchester	Study advisor: contributed to study design and reviewing the literature

Table 77. Chapter 10 acknowledgements. Conference presentations arising from the RTF exploratory trial workstream

Chapter 10: Conference presentations

Presenter	Title	Location and Date
Christine Wright	Feasibility and acceptability of a real time feedback intervention to improve patient experience in general practice: preliminary results	Society of Academic Primary Care Annual Scientific Meeting, Glasgow, July 2014
Mary Carter	Using touch screens in GP surgery waiting areas to collect real-time patient feedback: practice staff views of feasibility and acceptability	Society of Academic Primary Care South West meeting, March 2015

Acknowledgements

Chapter 11 - The validity and use of patient experience survey data in out-of-hours care

The out-of-hours workstream involved contributions from a wide variety of organisations and people. We would particularly like to thank all the out-of-hours service providers and their staff who participated in the research. We also thank all those service users who took the time to respond to the questionnaire they received. We are grateful to Ipsos MORI for their assistance in providing the GPPS data for analysis; Jonathan Jackson from the Health and Social Care Information Centre for providing the data to map out-of-hours GP providers to associated practices and for analytical input; and James Wallis from NHS England for analytical input. Finally, we thank Martin Roberts for statistical assistance.

Table 78. Chapter 11 acknowledgements. Contributions to the out-of-hours workstream

Chapter 11: Contributors

Name and role		Nature of contribution
John Campbell	Professor of General Practice & Primary Care University of Exeter	Co-Chief Investigator and Principal Investigator for this work stream. Oversaw the design and conduct of the study, the analysis and interpretation of the data, and edited the report as necessary
Martin Roland	Professor of Health Services Research University of Cambridge	Co-Chief Investigator: oversaw the design and conduct of the study, the analysis and interpretation of the data, and edited the report as necessary
Suzanne Richards	Senior Lecturer in Primary Care University of Exeter	Workstream lead: responsible for protocol design, study organisation, data collection, analysis and interpretation of the findings, drafting the report and critiquing all outputs for important intellectual content
Heather Barry	Associate Research Fellow University of Exeter	Study researcher: contributed to study design, organisation, data collection, and analysis and interpretation of the findings, and drafting the report.
Luke Mounce	Associate Research Fellow University of Exeter	Study researcher: contributed to study design, organisation, data collection, and analysis and interpretation of the findings, and drafting the report.
Antoinette Davey	Research Fellow University of Exeter	Study researcher: contributed to the data collection and data entry
William Henley	Professor of Medical Statistics	Study contributor: assisted with data analysis and interpretation

	University of Exeter	
Raff Calitri	Research Fellow University of Exeter	assisted with organisation, data collection, data analysis and interpretation
Fiona Warren	Lecturer in Medical Statistics, University of Exeter	Study contributor: assisted with data analysis and interpretation
Anthea Asprey	Associate Research Fellow University of Exeter	Study researcher: secondary analysis of qualitative data

Table 79. Chapter 11 acknowledgements. Conference presentations arising from the out-of-hours workstream

Chapter 11: Conference presentations

Presenter	Title	Location and Date
Heather Barry	Understanding patient experience of out-of-hours primary care in England	Society for Academic Primary Care South West meeting, March 2013
Heather Barry	Understanding patient experience of out-of-hours primary care: A pilot study	Society of Academic Primary Care Annual Scientific Meeting, Nottingham, July 2013
Heather Barry	A cross-sectional survey study of service users' experiences of out-of-hours primary medical care in England	Society for Academic Primary Care South West meeting, March 2014
Heather Barry	A cross-sectional survey study of service users' experiences of out-of-hours primary medical care in England	Society of Academic Primary Care Annual Scientific Meeting, Glasgow, July 2014
Luke Mounce	Establishing the potential validity of English GP Patient Survey items evaluating out-of-hours care	Society of Academic Primary Care South West meeting, March 2015
Suzanne Richards	How do staff from GP out-of-hours services use patient feedback to drive quality improvement? A qualitative interview study (POSTER)	Health Services Research Network, Annual Scientific Meeting, Nottingham, July 2015
Suzanne Richards	How do staff from GP out-of-hours services use patient feedback to drive quality improvement? A qualitative interview study (POSTER)	Society of Academic Primary Care Annual Scientific Meeting Oxford, July 2015

Table 80. Chapter 11 acknowledgements. Publications arising from the out-of-hours workstream

Chapter 11: Published papers		
Authors	**Title**	**Journal**
Warren FC, Abel G, Lyratzopoulos G, Elliott MN, Richards S, Barry HE et al.	Characteristics of service users and provider organisations associated with experience of out of hours general practitioner care in England: population based cross sectional postal questionnaire survey	*BMJ*. 2015;350:9

References

1. Institute of Medicine. *Crossing the quality chasm: a new health system for the 21st century*. Washington DC: National Academy Press; 1990.

2. Berwick DM, Nolan TW, Whittington J. The Triple Aim: Care, Health, And Cost. *Health Affairs* 2008;**27**:759-69.

3. Health Research Institute. *Scoring healthcare: Navigating customer experience ratings*. New York: PwC; 2013.

4. NHS England. *Improving general practice: a call to action.* 2013. URL: http://www.england.nhs.uk/ourwork/com-dev/igp-cta/ (accessed 3 May 2016).

5. Care Quality Commission Intelligent Monitoring. *NHS GP practices Indicators and methodology*. London; 2015. URL: https://www.cqc.org.uk/sites/default/files/20160105_Adult_Social_Care_Intelligent_Monitoring_Indicators_and_Methodology_for_publication.pdf. (accessed 12 April 2016).

6. Rozenblum R, Bates DW. Patient-centred healthcare, social media and the internet: the perfect storm? *BMJ Qual Saf* 2013;**22**(3):183-6.

7. HSCIC. *Quality and Outcomes Framework (QOF)*. URL: http://www.hscic.gov.uk/qof; 2014). (accessed 12 April 2016)

8. Campbell S, Reeves D, Kontopantelis E, Middleton E, Sibbald B, Roland M. Quality of Primary Care in England with the Introduction of Pay for Performance. *N Eng J Med* 2007;**357**:181-90.

9. Department of Health. *High Quality Care for All. NHS Next Stage Review Final Report.* London: Department of Health; 2008.

10. NHS England. *NHS Outcome Framework: Domain 4: Ensuring that people have a positive experience of care*. URL: http://www.england.nhs.uk/resources/resources-for-ccgs/out-frwrk/dom-4/ (accessed 3 May 2016).

11. University of Cambridge. *General Practice Assessment Questionnaire (GPAQ)*. URL: http://www.phpc.cam.ac.uk/gpaq/ (accessed 3 May 2016).

12. Health Foundation. *Quality improvement made simple: What everyone should know about healthcare quality improvement*. London: The Health Foundation; 2013.

13. Campbell SM, Reeves D, Kontopantelis E, Sibbald B, Roland M. Effects of Pay for Performance on the Quality of Primary Care in England. *N Engl J Med* 2009;**361**:368-78. http://dx.doi.org/doi:10.1056/NEJMsa0807651

14. Rodriguez HP, Rogers WH, Marshall RE, Safran DG. Multidisciplinary primary care teams: effects on the quality of clinician-patient interactions and organizational features of care. *Med Care* 2007;**45**:19-27. http://dx.doi.org/10.1097/01.mlr.0000241041.53804.29

15. Rodriguez HP, Marshall RE, Rogers WH, Safran DG. Primary Care Physician Visit Continuity: A Comparison of Patient-reported and Administratively Derived Measures. *J Gen Intern Med* 2008;**23**:1499-502. http://dx.doi.org/10.1007/s11606-008-0692-z

16. Bower P, Roland M, Campbell J, Mead N. Setting standards based on patients' views on access and continuity: secondary analysis of data from the general practice assessment survey. *BMJ* 2003;**326**:258. http://dx.doi.org/10.1136/bmj.326.7383.258

17. Sitzia J, Wood N. Patient satisfaction: a review of issues and concepts. *SocSciMed* 1997;**45**:1829-43.

18. Russell S. *Patients' experiences: top heavy with research. A literature review.* Melbourne, Australia: Bayside Medicare Local; 2013.

19. Salisbury C, Wallace M, Montgomery AA. Patients' experience and satisfaction in primary care: secondary analysis using multilevel modelling. *BMJ* 2010;**341**:c5004.

20. Buetow S. What do patients want from their general practitioners and their patients want from general practice and are they receiving it: a framework. *Soc Sci Med* 1995;**40**:213-21.

21. Wensing M, Jung HP, Mainz J, Olesen F, Grol R. A systematic review of the literature on patient priorities for general practice care. Part 1: Description of the research domain. *Soc Sci Med* 1998;**47**:1573-88.

22. Wofford MM, Wofford JL, Bothra J, Kendrick SB, Smith A, Lichstein PR. Patient complaints about physician behaviors: a qualitative study. *Acad Med* 2004;**79**:134-8.

23. Coulter A. *Trends in patients' experience of the NHS*. Oxford: Picker Institute Europe; 2006.

24. Benson J, Britten N. Patients' decisions about whether or not to take antihypertensive drugs: qualitative study. *BMJ* 2002;**325**:873. http://dx.doi.org/10.1136/bmj.325.7369.873

25. Griffith S. A review of the factors associated with patient compliance and the taking of prescribed medicines. *Br J Gen Pract* 1990;**40**:114-6.

26. Pound P, Britten N, Morgan M, Yardley L, Pope C, ker-White G, *et al.* Resisting medicines: a synthesis of qualitative studies of medicine taking. *Soc Sci Med* 2005;**61**:133-55.

27. Kim SS, Park BK. Patient-perceived communication styles of physicians in rehabilitation: the effect on patient satisfaction and compliance in Korea. *Am J Phys Med Rehabil* 2008;**87**:998-1005. http://dx.doi.org/10.1097/PHM.0b013e318186babf

28. Stewart M, Brown JB, Donner A, McWhinney IR, Oates J, Weston WW, *et al.* The impact of patient-centered care on outcomes. *J Fam Pract* 2000;**49**:796-804.

29. Orth JE, Stiles WB, Scherwitz L, Hennrikus D, Vallbona C. Patient exposition and provider explanation in routine interviews and hypertensive patients' blood pressure control. *Health Psychol* 1987;**6**:29-42.

30. Kuzel AJ, Woolf SH, Gilchrist VJ, Engel JD, LaVeist TA, Vincent C, *et al.* Patient reports of preventable problems and harms in primary health care. *Ann Fam Med* 2004;**2**:333-40.

31. Vincent C, Davy C, Esmail A, Neale G, Elstein M, Cozens JF, *et al.* Learning from litigation. The role of claims analysis in patient safety. *J Eval Clin Pract* 2006;**12**:665-74. http://dx.doi.org/10.1111/j.1365-2753.2006.00634.x

32. Vincent CA, Coulter A. Patient safety: what about the patient? *Qual Saf Health Care* 2002;**11**:76-80. http://dx.doi.org/10.1136/qhc.11.1.76

33. Koutantji M, Davis, R., Vincent, C. A., & Coulter, A. The patient's role in patient safety: engaging patients, their representatives, and health professionals. *Clinical Risk* 2005;**11**:99-104.

34. Macdonald W, Rogers A, Blakeman T, Bower P. Practice nurses and the facilitation of self-management in primary care. *J Adv Nurs* 2008;**62**:191-9. http://dx.doi.org/10.1111/j.1365-2648.2007.04585.x

35. Doyle C, Lennox L, Bell D. A systematic review of evidence on the links between patient experience and clinical safety and effectiveness. *BMJ Open* 2013;**3**. http://dx.doi.org/10.1136/bmjopen-2012-001570

36. Kennedy A, Nelson E, Reeves D, Richardson G, Roberts C, Robinson A, *et al.* A randomised controlled trial to assess the impact of a package comprising a patient-orientated, evidence-based self-help guidebook and patient-centred consultations on disease management and satisfaction in inflammatory bowel disease. *Health Technol Assess* 2003;**7**:iii, 1-113.

37. Zolnierek KB, Dimatteo MR. Physician communication and patient adherence to treatment: a meta-analysis. *Med Care* 2009;**47**:826-34. http://dx.doi.org/10.1097/MLR.0b013e31819a5acc

38. Carcaise-Edinboro P, Bradley CJ. Influence of patient-provider communication on colorectal cancer screening. *Med Care* 2008;**46**:738-45. http://dx.doi.org/10.1097/MLR.0b013e318178935a

39. Hsiao CJ, Boult C. Effects of quality on outcomes in primary care: a review of the literature. *Am J Med Qual* 2008;**23**:302-10. http://dx.doi.org/10.1177/1062860608315643

40. Patient Experience. *Understanding the Patient Experience.* URL: http://www.patientexperience.co.uk/en/understanding-the-patient-experience/ (Accessed 22 July 2015).

41. Robert G, Cornwell J. Rethinking policy approaches to measuring and improving patient experience. *J Health Serv Res Pol* 2013;**18**:67-9. http://dx.doi.org/10.1177/1355819612473583

42. Coulter A, Locock L, Ziebland S, Calabrese J. Collecting data on patient experience is not enough: they must be used to improve care. *BMJ* 2014;**348**:g2225. http://dx.doi.org/10.1136/bmj.g2225

43. Locock L, Robert G, Boaz A, Vougioukalou S, Shuldham C, Fielden J, *et al.* Using a national archive of patient experience narratives to promote local patient-centered quality improvement: an ethnographic process evaluation of 'accelerated' experience-based co-design. *J Health Serv Res Policy* 2014;**19**:200-7. http://dx.doi.org/10.1177/1355819614531565

44. NHS Practice Management Network. *A best practice guide to using real-time patient feedback*: NHS Practice Management Network; 2010.

45. Anderson JG, S., Saunders, K., Davis, A. . *Working to improve patient experience: real-time monitoring*. London: Imperial College Healthcare Trust, 2011. URL: www.ncin.org.uk/view?rid=1381 (accessed 3 May 2016)

46. Airey C, Bruster, S., Erens, B., Lilley, S. , Pickering, K., & Pitson, L. *National surveys of NHS patients General practice 1998*. London: Picker, Europe; 1999.

47. CFEP UK Surveys o. *Improving Practice Questionnaire (IPQ)*. URL: https://www.cfepsurveys.co.uk/OrganisationalAndPersonalDevelopment/PrimaryCare/IPQ (accessed 3 May 2016).

48. Roland M. Linking Physicians' Pay to the Quality of Care — A Major Experiment in the United Kingdom. *N Engl J Med* 2004;**351**:1448-54. http://dx.doi.org/doi:10.1056/NEJMhpr041294

49. Sainsbury R, Ditch J, Hutton S. Computer assisted personal interviewing. *Social Research Update* 1993;**3**:1-12.

50. Picker Institute Europe. *About NHS patient surveys*. URL: http://www.nhssurveys.org/ (accessed 3 May 2016).

51. Cleary PD, Crofton C, Hays RD, Horner R. Advances from the Consumer Assessment of Healthcare Providers and Systems (CAHPS(R)) Project. Introduction. *Med Care* 2012;**50 Suppl**:S1. http://dx.doi.org/10.1097/MLR.0b013e31826ec0cb

52. NHS Patient Survey Programme. *National surveys of NHS patients, 1998*. URL: http://webarchive.nationalarchives.gov.uk/20130107105354/http://dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_4035999.pdf (accessed 3 May 2016)

53. Hall JA, Dornan MC. What patients like about their medical care and how often they are asked: a meta-analysis of the satisfaction literature. *Soc Sci Med* 1988;**27**:935-9.

54. Cheraghi-Sohi S, Bower P, Mead N, McDonald R, Whalley D, Roland M. What are the key attributes of primary care for patients? Building a conceptual 'map' of patient preferences. *Health Expect* 2006;**9**:275-84. http://dx.doi.org/10.1111/j.1369-7625.2006.00395.x

55. Department of Health. *NHS Patient Experience Framework*. London: Department of Health, 2012. URL: https://www.gov.uk/government/publications/nhs-patient-experience-framework (accessed 3 May 2016)

56. NHS England *Review of the Friends and Family Test*. London: Department of Health, 2014. URL: http://www.england.nhs.uk/wp-content/uploads/2014/07/fft-rev1.pdf 2014 (accessed 3 May 2016)

57. NHS England. *Friends and Family Test in General Guidance: Data submission guidance*. London: NHS Employers, 2014. URL: https://www.england.nhs.uk/wp-content/uploads/2015/11/fft-gp-data-sub-nov15.pdf (accessed 3 May 2016)

58. Kmietowicz Z. Friends and family test “unfit” for comparing NHS services, finds research. *BMJ* 2014;**348**. http://dx.doi.org/10.1136/bmj.g4355

59. Sizmur S, Graham C, Walsh J. Influence of patients' age and sex and the mode of administration on results from the NHS Friends and Family Test of patient experience. *J Health Serv Res Policy* 2014;**20**:5-10.

60. Department of Health. *Raising standards for patients New partnerships in out-of-hours care; an independent review of GP out-of-hours services in England*. London: Department of Health; 2000.

61. House of Commons Health Committee. *GP Out-of-Hours Services Fifth Report of Session 2003–04*. URL: http://www.publications.parliament.uk/pa/cm200304/cmselect/cmhealth/697/69702.htm (accessed 3 May 2016).

62. National Audit Office. *The Provision of Out-of-Hours Care in England*. London: The Stationery Office, 2006.

63. UK Department of Health. *National quality requirements in the delivery of out-of-hours services,* 2004. URL: http://webarchive.nationalarchives.gov.uk/20080313140813/http://dh.gov.uk/en/publicationsandstatistics/publications/publicationspolicyandguidance/dh_4091213 (accessed 3 May 2016).

64. National Audit Office. *The Provision of Out-of-Hours Care in England*. URL: http://www.nao.org.uk/report/the-provision-of-out-of-hours-care-in-england/ (accessed 3 May 2016).

65. Hill JJ, Asprey A, Richards SH, Campbell JL. Multisource feedback questionnaires in appraisal and for revalidation: a qualitative study in UK general practice. *Br J Gen Pract* 2012;**62**:e314-21. http://dx.doi.org/10.3399/bjgp12X641429

66. IPSOS MORI. *GP Patient Survey Access Survey Final Technical Report 2007*. URL: https://gp-patient.co.uk/surveys-and-reports#june-2007 (accessed 3 May 2016)

67. Howie JG, Heaney DJ, Maxwell M, Walker JJ, Freeman GK. Developing a 'consultation quality index' (CQI) for use in general practice. *Fam Pract* 2000;**17**:455-61.

68. Mead N, Roland M. Understanding why some ethnic minority patients evaluate medical care more negatively than white patients: a cross sectional analysis of a routine patient survey in English general practices. *BMJ* 2009;**339**:b3450-b. http://dx.doi.org/10.1136/bmj.b3450

69. Kontopantelis E, Roland M, Reeves D. Patient experience of access to primary care: identification of predictors in a national patient survey. *BMC Fam Pract* 2010;**11**:61.

70. Ahmed F, Burt J, Roland M. Measuring Patient Experience: Concepts and Methods. *Patient* 2014; 10.1007/s40271-014-0060-5. http://dx.doi.org/10.1007/s40271-014-0060-5

71. Taira DA, Safran DG, Seto TB, Rogers WH, Inui TS, Montgomery J, *et al*. Do patient assessments of primary care differ by patient ethnicity? *Health Serv Res* 2001;**36**:1059-71.

72. Murray-Garcia JL, Selby JV, Schmittdiel J, Grumbach K, Quesenberry CP, Jr. Racial and ethnic differences in a patient survey: patients' values, ratings, and reports regarding physician primary care performance in a large health maintenance organization. *Med Care* 2000;**38**:300-10.

73. Elliott M, Haviland A, Kanouse D, Hambarsoomians K, Hays R. Adjusting for subgroup differences in extreme response tendency in ratings of health care: impact on disparity estimates. *Health Serv Res* 2009;**44**:542-61. http://dx.doi.org/ doi: 10.1111/j.1475-6773.2008.00922.x.

74. Weinick RM, Elliott MN, Volandes AE, Lopez L, Burkhart Q, Schlesinger M. Using standardized encounters to understand reported racial/ethnic disparities in patient experiences with care. *Health Serv Res* 2011;**46**:491-509. http://dx.doi.org/10.1111/j.1475-6773.2010.01214.x

75. King G, Murray CJ, Salomon JA, Tandon A. Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Polit Sci Rev* 2004;**98**:191-207.

76. Cheraghi-Sohi S, Bower P. Can the feedback of patient assessments, brief training, or their combination, improve the interpersonal skills of primary care physicians? A systematic review. *BMC Health Serv Res* 2008;**8**:179. http://dx.doi.org/10.1186/1472-6963-8-179

77. Berwick DM, James B, Coye MJ. Connections between quality measurement and improvement. *Med Care* 2003;**41**:I30-8.

78. Fung CH, Lim YW, Mattke S, Damberg C, Shekelle PG. Systematic review: the evidence that publishing patient care performance data improves quality of care. *Ann Intern Med* 2008;**148**:111-23.

79. Contandriopoulos D, Champagne F, Denis J-L. The multiple causal pathways between performance measures’ use and effects. *Med Care Res Rev* 2014;**71**:3-20.

80. Canadian Broadcasting Corporation (CBC). *Rate your hospital: online tool a 1st for Canadians,* 2013. URL: http://www.cbc.ca/news/health/rate-your-hospital-online-tool-a-1st-for-canadians-1.1336081 (accessed 3 May 2016)

81. Patient Opinion Australia. *Patient Opinion*. URL: https://www.patientopinion.org.au (accessed 3 May 2016)

82. The Leapfrog Group. *Compare Hospitals Now*. URL: http://www.leapfroggroup.org/cp (accessed 3 May 2016).

83. Richards SH, Pound P, Dickens A, Greco M, Campbell JL. Exploring users' experiences of accessing out-of-hours primary medical care services. *Qual Saf Health Care* 2007;**16**:469-77.

84. Campbell J, Roland M, Richards S, Dickens A, Greco M, Bower P. Users' reports and evaluations of out-of-hours health care and the UK national quality requirements: a cross sectional study. *Br J Gen Pract* 2009;**59**:e8-15. http://dx.doi.org/10.3399/bjgp09X394815

85. Department of Health. *Understanding what matters: A guide to using patient feedback to transform services*. Leeds: Department of Health; 2009.

86. Robert Wood Johnston Foundation. *Developing a Public Report for the CAHPS® Clinician & Group Survey: A Decision Guide*. 2013. URL: http://www.rwjf.org/content/dam/farm/reports/issue_briefs/2013/rwjf407929 (accessed 3 May 2016).

87. University Hospital of South Manchester. *UHSM launches pioneering transparency system*. Available at: http://www.uhsm.nhs.uk/news/Pages/transp.aspx (accessed 3 May 2016).

88. Collins K, O'Cathain A. The continuum of patient satisfaction--from satisfied to very satisfied. *Soc Sci Med* 2003;**57**:2465-70.

89. Edwards C, Staniszewska S, Crichton N. Investigation of the ways in which patients' reports of their satisfaction with healthcare are constructed. *Sociol Health Ill* 2004;**26**:159-83. http://dx.doi.org/10.1111/j.1467-9566.2004.00385.x

90. Williams B, Coyle J, Healy D. The meaning of patient satisfaction: an explanation of high reported levels. *Soc Sci Med* 1998;**47**:1351-9.

91. Lupton D, Donaldson C, Lloyd P. Caveat emptor or blissful ignorance? Patients and the consumerist ethos. *Soc Sci Med* 1991;**33**:559-68.

92. Lupton D. Consumerism, reflexivity and the medical encounter. *Soc Sci Med* 1997;**45**:373-81.

93. Henry SG, Fetters MD. Video elicitation interviews: a qualitative research method for investigating physician-patient interactions. *Ann Fam Med* 2012;**10**:118-25.

94. Paddison CAM, Abel GA, Roland MO, Elliott MN, Lyratzopoulos G, Campbell JL. Drivers of overall satisfaction with primary care: evidence from the English General Practice Patient Survey. *Health Expect* 2013; 10.1111/hex.12081. http://dx.doi.org/10.1111/hex.12081

95. Lofland J, Snow D, Anderson L, Lofland LH. *Analyzing Social Settings : A guide to qualitative observation and analysis*. 4th Edition. Belmont: Wadsworth; 2006.

96. Department of Health. *The NHS Plan: A Plan for Investment, A Plan for Reform*. London; 2000.

97. Goodyear-Smith F. Power Issues in the Doctor-Patient Relationship. *Health Care Analysis* 2001;**9**:449-62.

98. Pilnick A DR. On the remarkable persistence of asymmetry in doctor/patient interaction: a critical review. *Soc Sci Med* 2011;**72**:1374-82.

99. Kravitz RL, Callahan EJ, Paterniti D, Antonius D, Dunham M, Lewis CE. Prevalence and sources of patients' unmet expectations for care. *Ann Intern Med* 1996;**125**:730-7.

100. Marple RL, Kroenke K, Lucey CR, Wilder J, Lucas CA. Concerns and expectations in patients presenting with physical complaints: frequency, physician perceptions and actions, and 2-week outcome. *Arch Intern Med* 1997;**157**:1482-8.

101. Coyle J. Exploring the meaning of ‘dissatisfaction’with health care: the importance of ‘personal identity threat’. *Sociol Health Ill* 1999;**21**:95-123.

102. Campbell JL, Richards SH, Dickens A, Greco M, Narayanan A, Brearley S. Assessing the professional performance of UK doctors: an evaluation of the utility of the General Medical Council patient and colleague questionnaires. *Qual Saf Health Care* 2008;**17**:187-93. http://dx.doi.org/10.1136/qshc.2007.024679

103. Campbell J WC. *GMC Multi-Source Feedback Questionnaires. Interpreting and handling multisource feedback results: Guidance for appraisers*. Exeter: Peninsula College of Medicine and Dentistry; 2012. URL: http://dx.doi.org/http://www.gmc-uk.org/Information_for_appraisers.pdf_48212170.pdf (accessed 3 May 2016)

104. Elliott MN, Edwards C, Angeles J, Hambarsoomians K, Hays RD. Patterns of Unit and Item Nonresponse in the CAHPS® Hospital Survey. *Health Serv Res* 2005;**40**:2096-119. http://dx.doi.org/10.1111/j.1475-6773.2005.00476.x

105. Ipsos MORI. *GP Patient Survey - Technical Annex. 2013-14 Annual Report.* London: NHS England 2014. URL: http://gp-survey-

production.s3.amazonaws.com/archive/2014/July/1301375001_Technical%20Annex%202013-2014_FINAL%20v1.pdf (accessed 3 May 2016)

106. Klein DJ, Elliott MN, Haviland AM, Saliba D, Burkhart Q, Edwards C, *et al.* Understanding Nonresponse to the 2007 Medicare CAHPS Survey. *The Gerontologist* 2011;**51**:843-55. http://dx.doi.org/10.1093/geront/gnr046

107. Martino SC, Weinick RM, Kanouse DE, Brown JA, Haviland AM, Goldstein E, *et al.* Reporting CAHPS and HEDIS Data by Race/Ethnicity for Medicare Beneficiaries. *Health Serv Res* 2013;**48**:417-34. http://dx.doi.org/10.1111/j.1475-6773.2012.01452.x

108. O'Malley AJ, Zaslavsky AM, Elliott MN, Zaborski L, Cleary PD. Case-Mix Adjustment of the CAHPS® Hospital Survey. *Health Serv Res* 2005;**40**:2162-81. http://dx.doi.org/10.1111/j.1475-6773.2005.00470.x

109. McKinstry B, Walker J, Blaney D, Heaney D, Begg D. Do patients and expert doctors agree on the assessment of consultation skills? A comparison of two patient consultation assessment scales with the video component of the MRCGP. *Fam Pract* 2004;**21**:75-80.

110. Greco M, Spike N, Powell R, Brownlea A. Assessing communication skills of GP registrars: a comparison of patient and GP examiner ratings. *Med Educ* 2002;**36**:366-76.

111. Asprey A, Campbell JL, Newbould J, Cohn S, Carter M, Davey A, *et al.* Challenges to the credibility of patient feedback in primary healthcare settings: a qualitative study. *Br J Gen Pract* 2013;**63**:e200-e8.

112. Kenny DA, Veldhuijzen W, Weijden T, Leblanc A, Lockyer J, Legare F, *et al.* Interpersonal perception in the context of doctor-patient relationships: a dyadic analysis of doctor-patient communication. *Soc Sci Med* 2010;**70**:763-8. http://dx.doi.org/10.1016/j.socscimed.2009.10.065

113. Roberts MJ, Campbell JL, Richards SH, Wright C. Self-other agreement in multisource feedback: the influence of doctor and rater group characteristics. *J Contin Educ Health Prof* 2013;**33**:14-23. http://dx.doi.org/10.1002/chp.21162

114. Davis DA, Mazmanian PE, Fordis M, Van Harrison R, Thorpe KE, Perrier L. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *JAMA* 2006;**296**:1094-102. http://dx.doi.org/10.1001/jama.296.9.1094

115. Eva KW, Regehr G. "I'll never play professional football" and other fallacies of self-assessment. *J Contin Educ Health Prof* 2008;**28**:14-9. http://dx.doi.org/10.1002/chp.150

116. Gordon MJ. A review of the validity and accuracy of self-assessments in health professions training. *Acad Med* 1991;**66**:762-9.

117. Kruger J, Dunning D. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *J Pers Soc Psychol* 1999;**77**:1121-34.

118. Roberts MJ, Campbell JL, Abel GA, Davey AF, Elmore NL, Maramba I, *et al.* Understanding high and low patient experience scores in primary care: analysis of patients' survey data for general practices and individual doctors. *BMJ* 2014;**349**. http://dx.doi.org/10.1136/bmj.g6034

119. Campbell J, Smith P, Nissen S, Bower P, Elliott M, Roland M. The GP Patient Survey for use in primary care in the National Health Service in the UK – development and psychometric characteristics. *BMC Fam Pract* 2009;**10**:1-8. http://dx.doi.org/10.1186/1471-2296-10-57

120. Lyratzopoulos GE, M.; Barbiere, J. M.; Henderson, A.; Staetsky, L.; Paddison, C.; Campbell, J.; Roland, M. Understanding ethnic and other socio-demographic differences in patient experience of primary care: evidence from the English General Practice Patient Survey. *BMJ Qual Saf* 2012;**21**:21-9. http://dx.doi.org/10.1136/bmjqs-2011-000088

121. Burt J, Abel G, Elmore N, Campbell J, Roland M, Benson J, *et al.* Assessing the communication quality of consultations in primary care: initial reliability assessment of the Global Consultation Rating Scale, based on the Calgary-Cambridge Guide to the Medical Interview. *BMJ Open* 2014;**4**. http://dx.doi.org/doi:10.1136/bmjopen-2013-004339

122. Kurtz SM, Silverman JD. The Calgary-Cambridge Referenced Observation Guides: an aid to defining the curriculum and organizing the teaching in communication training programmes. *Med Educ* 1996;**30**:83-9.

123. Kurtz S, Silverman J, Benson J, Draper J. Marrying content and process in clinical method teaching: enhancing the Calgary-Cambridge guides. *Acad Med* 2003;**78**:802-9.

124. Rodriguez HP, Crane PK. Examining Multiple Sources of Differential Item Functioning on the Clinician & Group CAHPS® Survey. *Health Serv Res* 2011;**46**:1778-802. http://dx.doi.org/10.1111/j.1475-6773.2011.01299.x

125. Elliott MN, Zaslavsky AM, Goldstein E, Lehrman WG, Hambarsoomians K, Beckett MK, *et al.* Effects of survey mode, patient mix, and nonresponse on CAHPS® hospital survey scores *Health Serv Res* 2009;**44**:501-18. http://dx.doi.org/10.1111/j.1475-6773.2008.00914.x

126. Ipsos MORI. *GP Patient Survey National Summary Report: July 2015*. London: NHS England/Ipsos MORI. URL: http://gp-survey-production.s3.amazonaws.com/archive/2015/July/July%202015%20National%20Summary%20Report.pdf (accessed 3 May 2016)

127. Davoll S, Kowalski C, Kuhr K, Ommen O, Ernstmann N, Pfaff H. "Tendency to excuse" and patient satisfaction of those suffering with breast cancer. *Int J Public Health* 2013;**58**:385-93. http://dx.doi.org/10.1007/s00038-012-0405-6

128. Mann K, Gordon J, MacLeod A. Reflection and reflective practice in health professions education: a systematic review. *Adv Health Sci Educ Theory Pract* 2009;**14**:595-621. http://dx.doi.org/10.1007/s10459-007-9090-2

129. GMC. *Colleague and Patient Feedback for Revalidation*. London: General Medical Council; 2015. URL: http://www.gmc-uk.org/doctors/revalidation/colleague_patient_feedback.asp (accessed 3 May 2016)

130. Gillard S, Benson J, Silverman J. Teaching and assessment of explanation and planning in medical schools in the United Kingdom: cross sectional questionnaire survey. *Med Teach* 2009;**31**:328-31. http://dx.doi.org/10.1080/01421590801953018

131. Essers G, van Dulmen S, van Es J, van Weel C, van der Vleuten C, Kramer A. Context factors in consultations of general practitioner trainees and their impact on communication assessment in the authentic setting. *Patient Educ Couns* 2013;**93**:567-72. http://dx.doi.org/10.1016/j.pec.2013.08.024

132. Essers G, Dielissen P, van Weel C, van der Vleuten C, van Dulmen S, Kramer A. How do trained raters take context factors into account when assessing GP trainee communication performance? An exploratory, qualitative study. *Adv Health Sci Educ Theory Pract* 2015;**20**:131-47. http://dx.doi.org/10.1007/s10459-014-9511-y

133. Lyratzopoulos G, Elliott MN, Barbiere JM, Staetsky L, Paddison CA, Campbell J, *et al.* How can health care organizations be reliably compared?: Lessons from a national survey of patient experience. *Med Care* 2011;**49**:724-33. http://dx.doi.org/10.1097/MLR.0b013e31821b3482

134. Aboulghate A, Abel G, Elliott MN, Parker RA, Campbell J, Lyratzopoulos G, *et al.* Do English patients want continuity of care, and do they receive it? *Br J Gen Pract* 2012;**62**:e567-75. http://dx.doi.org/10.3399/bjgp12X653624

135. Freeman GK. *Continuity of Care Toolkit: helping clinicians and practices maximise relationship continuity*. London: RCGP; 2013.

136. Roland M, Paddison C. Better management of patients with multimorbidity. *BMJ* 2013;**2013**:f2510.

137. Campbell J, Ramsay J, Green J. Age, gender, socioeconomic, and ethnic differences in patients' assessments of primary health care. *Quality in Health Care : QHC* 2001;**10**:90-5. http://dx.doi.org/10.1136/qhc.10.2.90

138. Venn S, Fone DL. Assessing the influence of socio-demographic factors and health status on expression of satisfaction with GP services. *Clin Govern Int J* 2005;**10**:118-25. http://dx.doi.org/doi:10.1108/14777270510594290

139. Croker JE, Swancutt DR, Roberts MJ, Abel GA, Roland M, Campbell JL. Factors affecting patients' trust and confidence in GPs: evidence from the English national GP patient survey. *BMJ Open* 2013;**3**. http://dx.doi.org/10.1136/bmjopen-2013-002762

140. Weech-Maldonado R, Elliott MN, Adams JL, Haviland AM, Klein DJ, Hambarsoomian K, *et al.* Do Racial/Ethnic Disparities in Quality and Patient Experience within Medicare Plans Generalize across Measures and Racial/Ethnic Groups? *Health Serv Res* 2015; 10.1111/1475-6773.12297. http://dx.doi.org/10.1111/1475-6773.12297

141. Anhang Price R, Haviland AM, Hambarsoomian K, Dembosky JW, Gaillot S, Weech-Maldonado R, *et al.* Do Experiences with Medicare Managed Care Vary According to the Proportion of Same-Race/Ethnicity/Language Individuals Enrolled in One's Contract? *Health Serv Res* 2015; 10.1111/1475-6773.12292. http://dx.doi.org/10.1111/1475-6773.12292

142. Goldstein E, Elliott MN, Lehrman WG, Hambarsoomian K, Giordano LA. Racial/Ethnic Differences in Patients’ Perceptions of Inpatient Care Using the HCAHPS Survey. *Med Care Res Rev* 2010;**67**:74-92. http://dx.doi.org/10.1177/1077558709341066

143. Weech-Maldonado R, Morales LS, Elliott M, Spritzer K, Marshall G, Hays RD. Race/Ethnicity, Language, and Patients' Assessments of Care in Medicaid Managed Care. *Health Serv Res* 2003;**38**:789-808. http://dx.doi.org/10.1111/1475-6773.00147

144. Weinick R, Haviland A, Hambarsoomian K, Elliott MN. Does the Racial/Ethnic Composition of Medicare Advantage Plans Reflect Their Areas of Operation? *Health Serv Res* 2014;**49**:526-45. http://dx.doi.org/10.1111/1475-6773.12100

145. Department for Communities and Local Government. *The English Indices of Deprivation 2010*. London; 2010. URL: https://www.gov.uk/government/statistics/english-indices-of-deprivation-2010 (accessed 3 May 2016)

146. Lyratzopoulos G, Neal RD, Barbiere JM, Rubin GP, Abel GA. Variation in number of general practitioner consultations before hospital referral for cancer: findings from the 2010 National Cancer Patient Experience Survey in England. *Lancet Oncol* 2012;**13**:353-65. http://dx.doi.org/10.1016/s1470-2045(12)70041-4

147. EuroQol Group. EuroQol--a new facility for the measurement of health-related quality of life. *Health Pol* 1990;**16**:199-208.

148. Rasch G. *Probabilistic Models for Some Intelligence and Attainment Tests Copenhagen*. Denmark: Danmarks Paedogogiske Institut; 1960.

149. Hambleton R SH, Rogers H. *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage Press; 1991.

150. Lord FM. *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ: Lawrence Erlbaum; 1980.

151. Thissen D SL, Wainer H. Detection of differential item functioning using the parameters of item response models. In: Holland P, Wainer, H., editor. *Differential Item Functioning*Hillsdale, NJ: Erlbaum; 1993:67–111.

152. Holland P. A framework and history for score linking. In: Dorans N PM, Holland P., editor. *Linking and Aligning Scores and Scales*New York: NY: Springer; 2007:5–30.

153. Kopf J, Zeileis A, Strobl C. A Framework for Anchor Methods and an Iterative Forward Approach for DIF Detection. *Applied Psychological Measurement* 2015;**39**:83-103. http://dx.doi.org/10.1177/0146621614544195

154. Cai L. flexMIRT: Flexible multilevel item factor analysis and test scoring [Computer software]. In. Seattle, WA: Vector Psychometric Group, LLC; 2012.

155. Cudeck R, Henly SJ. Model selection in covariance structures analysis and the "problem" of sample size: a clarification. *Psychol Bull* 1991;**109**:512-9.

156. Hays RD, Morales LS, Reise SP. Item response theory and health outcomes measurement in the 21st century. *Med Care* 2000;**38**:Ii28-42.

157. Public Health England. *Metadata for National General Practice Profiles*. URL: http://www.apho.org.uk/resource/item.aspx?RID=95729 (accessed 3 May 2016).

158. Ahmed F, Abel GA, Lloyd CE, Burt J, Roland M. Does the availability of a South Asian language in practices improve reports of doctor-patient communication from South Asian patients? Cross sectional analysis of a national patient survey in English general practices. *BMC Fam Pract* 2015;**16**:55.

159. Office for National Statistics. *Language in England and Wales, 2011.* London; 2014. URL: http://www.ons.gov.uk/ons/rel/census/2011-census-analysis/language-in-england-and-wales-2011/rpt---language-in-england-and-wales--2011.html#tab-Proficiency-in-English. (accessed 3 May 2016)

160. Ahmed N. Language, Gender and Citizenship: Obstacles in the Path to Learning English for Bangladeshi Women in London's East End. *Sociological Research Online* 2008;**13**:12.

161. Meischke HW, Calhoun RE, Yip MP, Tu SP, Painter IS. The effect of language barriers on dispatching EMS response. *Prehosp Emerg Care* 2013;**17**:475-80. http://dx.doi.org/10.3109/10903127.2013.811565

162. John-Baptiste A, Naglie G, Tomlinson G, Alibhai SMH, Etchells E, Cheung A, *et al.* The Effect of English Language Proficiency on Length of Stay and In-hospital Mortality. *J Gen Intern Med* 2004;**19**:221-8. http://dx.doi.org/10.1111/j.1525-1497.2004.21205.x

163. Sarver J, Baker DW. Effect of Language Barriers on Follow-up Appointments After an Emergency Department Visit. *J Gen Intern Med* 2000;**15**:256-64. http://dx.doi.org/10.1111/j.1525-1497.2000.06469.x

164. Parsons JA, Baker NA, Smith-Gorvie T, Hudak PL. To ‘Get by’ or ‘get help’? A qualitative study of physicians’ challenges and dilemmas when patients have limited English proficiency. *BMJ Open* 2014;**4**. http://dx.doi.org/10.1136/bmjopen-2013-004613

165. Hunt LM, Schneider S, Comer B. Should "acculturation" be a variable in health research? A critical review of research on US Hispanics. *Soc Sci Med* 2004;**59**:973-86. http://dx.doi.org/10.1016/j.socscimed.2003.12.009

166. Hasnain M, Schwartz A, Girotti J, Bixby A, Rivera L. Differences in patient-reported experiences of care by race and acculturation status. *J Immigr Minor Health* 2013;**15**:517-24. http://dx.doi.org/10.1007/s10903-012-9728-x

167. Lopez-Class M, Castro FG, Ramirez AG. Conceptions of acculturation: a review and statement of critical issues. *Soc Sci Med* 2011;**72**:1555-62. http://dx.doi.org/10.1016/j.socscimed.2011.03.011

168. Zambrana RE, Carter-Pokras O. Role of acculturation research in advancing science and practice in reducing health care disparities among Latinos. *Am J Public Health* 2010;**100**:18-23. http://dx.doi.org/10.2105/ajph.2008.138826

169. Smith WR, Betancourt JR, Wynia MK, Bussey-Jones J, Stone VE, Phillips CO, *et al.* Recommendations for teaching about racial and ethnic disparities in health and health care. *Ann Intern Med* 2007;**147**:654-65.

170. Betancourt JR. Cross-cultural medical education: conceptual approaches and frameworks for evaluation. *Acad Med* 2003;**78**:560-9.

171. Burgess DJ. Addressing Racial Healthcare Disparities: How Can We Shift the Focus from Patients to Providers? *J Gen Intern Med* 2011;**26**:828-30. http://dx.doi.org/10.1007/s11606-011-1748-z

172. Torres S. Expanding the gerontological imagination on ethnicity: conceptual and theoretical perspectives. *Ageing and Society* 2015;**35**:935-60.

173. Groves RM, Peytcheva E. The impact of nonreponse rates on nonresponse bias. *Public Opinion Quarterly* 2008;**72**:167-89. http://dx.doi.org/10.1093/poq/nfn011

174. Karliner LS, Jacobs EA, Chen AH, Mutha S. Do Professional Interpreters Improve Clinical Care for Patients with Limited English Proficiency? A Systematic Review of the Literature. *Health Serv Res* 2007;**42**:727-54. http://dx.doi.org/10.1111/j.1475-6773.2006.00629.x

175. Greenhalgh T. Health literacy: towards system level solutions. *BMJ* 2015;**350**:h1026. http://dx.doi.org/10.1136/bmj.h1026

176. Burt J, Lloyd C, Campbell J, Roland M, Abel G. Variations in GP-patient communication by ethnicity, age and gender: Evidence from a national patient survey. *BJGP* 2016;**66(642)**:e47-52.

177. Setodji CM, Elliott MN, Abel G, Burt J, Roland M, Campbell J. Evaluating Differential Item Functioning in the English General Practice Patient Survey: Comparison of South Asian and White British Subgroups. *Med Care* 2015; 10.1097/mlr.0000000000000397. http://dx.doi.org/10.1097/mlr.0000000000000397

178. Brady HE. The Perils of Survey Research: Inter-Personally Incomparable Responses. *Political Methodology* 1985;**11**:269-91. http://dx.doi.org/10.2307/41289344

179. Hillen MA, van Vliet LM, de Haes HCJM, Smets EMA. Developing and administering scripted video vignettes for experimental research of patient-provider

communication. *Patient Educ Couns* 2013;**91**:295-309. http://dx.doi.org/DOI 10.1016/j.pec.2013.01.020

180. Bécares L, Das-Munshi J. Ethnic density, health care seeking behaviour and expected discrimination from health services among ethnic minority people in England. *Health & Place* 2013;**22**:48-55. http://dx.doi.org/http://dx.doi.org/10.1016/j.healthplace.2013.03.005

181. van Vliet LM, van der Wall E, Albada A, Spreeuwenberg PM, Verheul W, Bensing JM. The validity of using analogue patients in practitioner-patient communication research: systematic review and meta-analysis. *J Gen Intern Med* 2012;**27**:1528-43. http://dx.doi.org/10.1007/s11606-012-2111-8

182. NHS England. *GP Patient Survey: Survey and Reports*. URL: https://gp-patient.co.uk/surveys-and-reports (accessed 3 May 2016).

183. Institute of Medicine. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*. Washington, DC: The National Academic Press; 2003.

184. Edwards A, Evans R, White P, Elwyn G. Experiencing patient-experience surveys: a qualitative study of the accounts of GPs. *Br J Gen Pract* 2011;**61**:e157-e66. http://dx.doi.org/10.3399/bjgp11X567072

185. Van Der Vleuten CPM. The assessment of professional competence: Developments, research and practical implications. *Adv Health Sci Ed* 1996;**1**:41-67. http://dx.doi.org/10.1007/BF00596229

186. Postgraduate Medical Education and Training Board (PMETB). *Developing and maintaining an assessment system - a PMETB guide to good practice*. London: PMETB; 2007. URL: https://www.researchgate.net/publication/264405740_Developing_and_Maintaining_an_Assessment_System-a_PMETB_Guide_to_Good_Practice (accessed 3 May 2016)

187. NHS England. *General Practice Patient Survey*. 2015. URL: https://gp-patient.co.uk/ (accessed 3 May 2016).

188. Ritchie J SL, Bryman A, Burgess RG. *Qualitative data analysis for applied policy research. Analyzing Qualitative Data.* London: Routledge; 1994.

189. Reinders ME, Ryan BL, Blankenstein AH, van der Horst HE, Stewart MA, van Marwijk HW. The effect of patient feedback on physicians' consultation skills: a systematic review. *Acad Med* 2011;**86**:1426-36. http://dx.doi.org/10.1097/ACM.0b013e3182312162

190. Sargeant J, Mann K, Ferrier S. Exploring family physicians' reactions to multisource feedback: perceptions of credibility and usefulness. *Med Educ* 2005;**39**:497-504. http://dx.doi.org/10.1111/j.1365-2929.2005.02124.x

191. Davies EA, Meterko MM, Charns MP, Nealon Seibert ME, Cleary PD. Factors affecting the use of patient survey data for quality improvement in the Veterans Health Administration. *BMC Health Serv Res* 2011;**11**:334-. http://dx.doi.org/10.1186/1472-6963-11-334

192. Vingerhoets E, Wensing M, Grol R. Feedback of patients' evaluations of general practice care: a randomised trial. *Qual Health Care* 2001;**10**:224-8.

193. Wensing M, Vingerhoets E, Grol R. Feedback based on patient evaluations: a tool for quality improvement? *Patient Educ Couns* 2003;**51**:149-53.

194. Reeves R, Seccombe I. Do patient surveys work? The influence of a national survey programme on local quality-improvement initiatives. *Qual Saf Health Care* 2008;**17**:437-41. http://dx.doi.org/DOI 10.1136/qshc.2007.022749

195. Boyer L, Francois P, Doutre E, Weil G, Labarere J. Perception and use of the results of patient satisfaction surveys by care providers in a French teaching hospital. *Int J Qual Health Care* 2006;**18**:359-64. http://dx.doi.org/10.1093/intqhc/mzl029

196. Reinders ME, Blankenstein AH, van der Horst HE, Knol DL, Schoonheim PL, van Marwijk HW. Does patient feedback improve the consultation skills of general practice trainees? A controlled trial. *Med Educ* 2010;**44**:156-64. http://dx.doi.org/10.1111/j.1365-2923.2009.03569.x

197. Reeves R, West E, Barron D. Facilitated patient experience feedback can improve nursing care: a pilot study for a phase III cluster randomised controlled trial. *BMC Health Serv Res* 2013;**13**:259. http://dx.doi.org/10.1186/1472-6963-13-259

198. Department of Health. *The NHS Outcomes Framework 2013-14*. London: Department of Health; 2012.

199. Fitzpatrick R, Graham. C., Gibbons, E., King, J., Flott, K. and Jenkinson, C. *Development of New Models for Collection and Use of Patient Experience Information in the NHS*. Oxford: Picker Institute Europe; 2014.

200. Williams TR. A Critique of Some Assumptions of Social Survey Research. *Public Opinion Quarterly* 1959;**23**:55-62. http://dx.doi.org/10.1086/266845

201. Elbaum B. Challenges in Interpreting Accountability Results for Schools' Facilitation of Parent Involvement Under IDEA. *Journal of Disability Policy Studies* 2014;**24**:206-17. http://dx.doi.org/10.1177/1044207312461947

202. Hibbard JH, Greene J, Daniel D. What Is Quality Anyway? Performance Reports That Clearly Communicate to Consumers the Meaning of Quality of Care. *Med Care Res Rev* 2010;**67**:275-93. http://dx.doi.org/10.1177/1077558709356300

203. Braun V, Clarke V. Using thematic analysis in psychology. *Qualitative Research in Psychology* 2006;**3**:77-101.

204. May C, Rapley T, Moreira T, Finch T, Heaven B. Technogovernance: evidence, subjectivity, and the clinical encounter in primary care medicine. *Soc Sci Med* 2006;**62**:1022-30. http://dx.doi.org/10.1016/j.socscimed.2005.07.003

205. Boyd DaM, E., eds. *Feedback in Higher and Professional Education: Understanding it and doing it well*. Abingdon: Routledge; 2013.

206. Iacobucci G. General practices could be subject to "Ofsted style" ratings within two years. *BMJ* 2013;**346**:f1941. http://dx.doi.org/10.1136/bmj.f1941

207. Wright C, Richards SH, Hill JJ, Roberts MJ, Norman GR, Greco M, *et al.* Multisource feedback in evaluating the performance of doctors: the example of the UK General Medical Council patient and colleague questionnaires. *Acad Med* 2012;**87**:1668-78. http://dx.doi.org/10.1097/ACM.0b013e3182724cc0

208. Roland M, Roberts M, Rhenius V, Campbell J. GPAQ-R. Development and psychometric properties of a version of the General Practice Assessment Questionnaire for use for revalidation by general practitioners in the UK. *BMC Fam Pract* 2013;**14**:160. http://dx.doi.org/10.1186/1471-2296-14-160

209. Lyratzopoulos G, Elliott M, Barbiere JM, Henderson A, Staetsky L, Paddison C, *et al.* Understanding ethnic and other socio-demographic differences in patient experience of primary care: evidence from the English General Practice Patient Survey. *BMJ Qual Saf* 2012;**21**:21-9. http://dx.doi.org/10.1136/bmjqs-2011-000088

210. Robinson GK. That BLUP is a Good Thing: The Estimation of Random Effects. *Stat Sci* 1991;**6**:15-32. http://dx.doi.org/10.1214/ss/1177011933

211. Downing SM. Reliability: on the reproducibility of assessment data. *Med Ed* 2004;**38**:1006-12.

212. Lockyer J. Multisource feedback in the assessment of physician competencies. *J Contin Educ Health Prof* 2003;**23**:4-12. http://dx.doi.org/10.1002/chp.1340230103

213. Holm S. A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics* 1979;**6**:65-70.

214. Cohen J. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychol Bull* 1968;**70**:213.

215. Safran DG, Karp M, Coltin K, Chang H, Li A, Ogren J, *et al.* Measuring Patients' Experiences with Individual Primary Care Physicians: Results of a Statewide Demonstration Project. *J Gen Intern Med* 2006;**21**:13-21. http://dx.doi.org/10.1111/j.1525-1497.2005.00311.x

216. Haggerty JL, Pineault R, Beaulieu MD, Brunelle Y, Gauthier J, Goulet F, *et al.* Practice features associated with patient-reported accessibility, continuity, and coordination of primary health care. *Ann Fam Med* 2008;**6**:116-23. http://dx.doi.org/10.1370/afm.802

217. Mazor KM, Clauser BE, Field T, Yood RA, Gurwitz JH. A demonstration of the impact of response bias on the results of patient satisfaction surveys. *Health Serv Res* 2002;**37**:1403-17.

218. Ramsay J, Campbell JL, Schroter S, Green J, Roland M. The General Practice Assessment Survey (GPAS): tests of data quality and measurement properties. *Fam Pract* 2000;**17**:372-9.

219. Haskard K, Williams S, DiMatteo M, Rosenthal R, White M, Goldstein M. Physician and patient communication training in primary care: effects on participation and satisfaction. *Health Psychol* 2008;**27**:513-22.

220. Byrne P, Long B. *Doctors Talking to Patients*. London: HMSO; 1976.

221. Checkland K. Understanding general practice: a conceptual framework developed from case studies in the UK NHS. . *Br J Gen Pract* 2007;**57**:56-63.

222. Armstrong D, Ogden J. The role of etiquette and experimentation in explaining how doctors change behaviour: a qualitative study. *Sociol Health Ill* 2006;**28**:951-68.

223. King EB, de Chermont K, West M, Dawson JF, Hebl MR. How innovation can alleviate negative consequences of demanding work contexts: The influence of climate for innovation on organizational outcomes. *J Occup Organ Psychol* 2007;**80**:631-45.

224. West MA, Borrill C. The influence of team working. In: Cox J, King J, Hutchinson A, McAvoy P, editors. *Understanding doctor's performance.* Oxford: Radcliffe Publishing; 2006:106-22.

225. Anderson NR, West MA. Measuring climate for work group innovation: development and validation of the team climate inventory. *J Organ Behav* 1998;**19**:235-58.

226. Anderson NR, West MA. The Team Climate Inventory: Development of the TCI and its application in teambuilding for innovativeness. *Eur J Work Organ Psy* 1996;**5**:53-66.

227. Bower P, Campbell S, Bojke C, Sibbald B. Team structure, team climate and the quality of care in primary care: an observational study. *Qual Saf Health Care* 2003;**12**:273-79.

228. Campbell SM, Hann M, Hacker J, Burns C, Oliver D, Thapar A, *et al.* Identifying predictors of high quality care in English general practice: observational study. *BMJ* 2001;**323**:1-6.

229. Proudfoot J, Jayasinghe UW, Holton C, Grimm J, Bubner T, Amoroso C, *et al.* Team climate for innovation: what difference does it make in general practice? *Int J Qual Health Care* 2007;**3**:164-9.

230. Kivimaki M, Vanhala A, Pentti J, Lansisalmi H, Virtanen M, Elovianio M, *et al.* Team climate, intention to leave and turnover among hospital employees: Prospective cohort study. *BMC Health Serv Res* 2007;**7**:170.

231. Goh TT, Eccles MP, Steen N. Factors predicting team climate and its relationship with quality of care in general practice. *BMC Health Serv Res* 2009;**9**:138.

232. Hann M, Bower P, Campbell S, Marshall M, Reeves D. The association between culture, climate and quality of care in primary health care teams. *Fam Pract* 2007;**24**:323-9.

233. Brown H, Davidson D, Ellins J. *NHS West Midlands Investing for Health: Real-time Patient Feedback Project – Final Report*. Birmingham: University of Birmingham and NHS West Midlands, Health Services Management Centre, 2009.

234. Dirocco DN, Day SC. Obtaining patient feedback at point of service using electronic kiosks. *Am J Manag Care* 2011;**17**:e270-6.

235. Greenhalgh T, Swinglehurst D. Studying technology use as social practice: the untapped potential of ethnography. *BMC Med* 2011;**9**:45.

236. Bentley ME, Boot, M.T., Gittelsohn, J., Stallings, R.Y. *The use of structured observations in the study of health behaviour (Occasional Paper 27).* The Hague,

Netherlands: IRC International Water and Sanitation Centre/London School of Hygiene and Tropical Medicine.

237. Srigley JA, Furness CD, Baker GR, Gardam M. Quantification of the Hawthorne effect in hand hygiene compliance monitoring using an electronic monitoring system: a retrospective cohort study. *BMJ Qual Saf* 2014;**23**:974-80.

238. Curtis L. *Unit Costs of Health and Social Care 2014*. Canterbury: University of Kent at Canterbury; 2015.

239. May C, Finch T. Implementing, embedding, and integrating practices: an outline of normalization process theory. *Sociology* 2009;**43**:535-54.

240. Wofford JL, Campos CL, Jones RE, Stevens SF. Real-time patient survey data during routine clinical activities for rapid-cycle quality improvement. *JMIR Med Inform* 2015;**3**:e13. http://dx.doi.org/10.2196/medinform.3697

241. Deming WE. *Out of the crisis*. Cambridge MA MIT Press; 2000.

242. Warren FC, Abel G, Lyratzopoulos G, Elliott MN, Richards S, Barry HE, et al. Characteristics of service users and provider organisations associated with experience of out of hours general practitioner care in England: population based cross sectional postal questionnaire survey. *BMJ* 2015;350:h2040

243. National Audit Office. *Out-of-hours GP services in England*. London: Report by the Controller and Auditor General; 2014.

244. Colin-Thomé D, Field S. *General Practice Out-of-Hours Services: Project to Consider and Assess Current Arrangements*. Care Quality Commission, London; 2010.

245. Healthcare Commission. *Not just a matter of time. A review of urgent and emergency care services in England*. London: Healthcare Commission; 2008.

246. Cosford PA, Thomas JM. Safer out of hours primary care. *BMJ* 2010;**340**:c3194.

247. Egbunike JN, Shaw C, Bale S, Elwyn G, Edwards A. Understanding patient experience of out-of-hours general practitioner services in South Wales: a qualitative study. *Emerg Med J* 2008;**25**:649-54. http://dx.doi.org/10.1136/emj.2007.052001

248. Poole R, Gamper A, Porter A, Egbunike J, Edwards A. Exploring patients' self-reported experiences of out-of-hours primary care and their suggestions for improvement: a qualitative study. *Fam Pract* 2011;**28**:210-9. http://dx.doi.org/10.1093/fampra/cmq090

249. O'Cathain A, Knowles E, Turner J, Nicholl J. Acceptability of NHS 111 the telephone service for urgent health care: cross sectional postal survey of users' views. *Fam Pract* 2014;**31**:193-200. http://dx.doi.org/10.1093/fampra/cmt078

250. Care Quality Commission. A fresh start for the regulation and inspection of GP practices and GP out-of-hours services. Newcastle upon Tyne: Care Quality Commission; 2013.

251. Care Quality Commission. *Our new approach to the inspection of NHS GP out-of-hours services: Findings from the first comprehensive inspections*; Newcastle Upon Tyne: Care Quality Commission, 2014.

252. NHS England. Primary medical care functions delegated to clinical commissioning groups: guidance. . London: NHS England; 2014.

253. Garratt A, Danielsen K, Hunskaar S. Patient satisfaction questionnaires for primary care out-of-hours services: a systematic review. *Br J Gen Pract* 2007;**57**:741-7.

254. Paddison C, Elliott M, Parker R, Staetsky L, Lyratzopoulos G, Campbell JL, *et al.* Should measures of patient experience in primary care be adjusted for case mix? Evidence from the English General Practice Patient Survey. *BMJ Qual Saf* 2012;**21**:634-40.

255. Overeem K, Faber MJ, Arah OA, Elwyn G, Lombarts KMJM, Wollersheim HC, *et al.* Doctor performance assessment in daily practise: does it help doctors or not? A systematic review. *Med Ed* 2007;**41**:1039-49.

256. Allemann Iseli M, Kunz R, Blozik E. Instruments to assess patient satisfaction after teleconsultation and triage: a systematic review. *Patient Prefer Adherence* 2014;**8**:893-907. http://dx.doi.org/10.2147/ppa.s56160

257. Office for National Statistics. *Quality and methodology information for population estimates by ethnic group (PEEG)*. 2011. URL: http://www.ons.gov.uk/ons/taxonomy/index.html?nscl=Population+Estimates+by+Ethnic+Group (accessed 3 May 2016).

258. McLennan D, Barnes H, Noble M, Davies J, Garratt E, Dibben C. *English Indices of Deprivation 2010*; 2011. URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/6320/1870718.pdf (accessed 3 May 2016)

259. Jha AK, Orav EJ, Zheng J, Epstein AM. Patients' perceptions of hospital care in the United States. *N Engl J Med* 2008;**359**:1921-31.

260. Lehrman W, Elliott M, Goldstein E, Beckett M, Klein D, Giordano L. Characteristics of Hospitals Demonstrating Superior Performance in Patient Experience and Clinical Process Measures of Care. *Med Care Res Rev* 2010;**67**:38-55.

261. McKinley RK, Stevenson K, Adams S, Manku-Scott TK. Meeting patient expectations of care: the major determinant of satisfaction with out-of-hours primary medical care? *Fam Pract* 2002;**19**:333-8.

262. Richards SH, Coast J, Peters TJ. Patient reported use of health service resources compared with information from health providers. *Health Soc Care Community* 2003;**11**(6):510-8

263. Groves RM. Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly* 2006;**70**:646-75. http://dx.doi.org/10.1093/poq/nfl033

264. van Uden C, Ament A, Hobma S, Zwietering P, Crebolder H. Patient satisfaction with out-of-hours primary care in the Netherlands. *BMC Health Serv Res* 2005;**5**:6.

265. Willis G. *Cognitive interviewing: a tool for improving questionnaire design*. Thousand Oaks, California, US; 2005.

266. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling* 1999;**6**:1-55.

267. Bentler PM. Comparative fit indexes in structural models. *Psychol Bull* 1990;**107**:238-46.

268. Russell DW. In search of underlying dimensions: The use (and abuse) of factor analysis in Personality and Social Psychology Bulletin. *Personality and Social Psychology Bulletin* 2002;**28**:1629–46.

269. Kolenikov S, Angeles G, editors. *The Use of Discrete Data in PCA: Theory, Simulations, and Applications to Socioeconomic Indices*. Chapel Hill, N.C.: Carolina Population Center MEASURE Evaluation, University of North Carolina at Chapel Hill; 2004.

270. Freeman GK, Horder JP, Howie JGR, Hungin AP, Hill AP, Shah NC, *et al*. Evolving general practice consultation in Britain: issues of length and context. *BMJ* 2002;**324**:880. http://dx.doi.org/10.1136/bmj.324.7342.880

271. Strauss A, Corbin J. *Basics of qualitative research: techniques and procedures for developing grounded theory*. 2nd edn. Newbury Park, CA: SAGE Publications 1998.

272. Riiskjaer E, Ammentorp J, Kofoed PE. The value of open-ended questions in surveys on patient experience: number of comments and perceived usefulness from a hospital

perspective. *Int J Qual Health Care* 2012;**24**:509-16. http://dx.doi.org/10.1093/intqhc/mzs039

273. The Health Foundation. *Measuring patient experience. Evidence scan.* London: The Health Foundation, 2013. URL: http://www.health.org.uk/sites/default/files/MeasuringPatientExperience.pdf (accessed 3 May 2016)

274. Jamtvedt G, Young JM, Kristoffersen DT, Thomson O'Brien MA, Oxman AD. Audit and feedback: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev* 2003; 10.1002/14651858.CD000259:CD000259. http://dx.doi.org/10.1002/14651858.CD000259

275. Nielsen J, Riiskjaer E. From patient surveys to organizational change: rational change processes and institutional forces. *Journal of Change Management* 2013;**13**:179-205.

276. Luxford K, Safran DG, Delbanco T. Promoting patient-centered care: a qualitative study of facilitators and barriers in healthcare organizations with a reputation for improving the patient experience. *Int J Qual Health Care* 2011;**23**:510-5. http://dx.doi.org/DOI 10.1093/intqhc/mzr024

277. Manary MP, Boulding W, Staelin R, Glickman SW. The Patient Experience and Health Outcomes. *N Engl J Med* 2013;**368**:201-3. http://dx.doi.org/Doi 10.1056/Nejmp1211775

278. Boiko O, Campbell JL, Elmore N, Davey AF, Roland M, Burt J. The role of patient experience surveys in quality assurance and improvement: a focus group study in English general practice. *Health Expect* 2014; 10.1111/hex.12298. http://dx.doi.org/10.1111/hex.12298

279. Llanwarne NR, Abel GA, Elliott MN, Paddison CA, Lyratzopoulos G, Campbell JL, *et al.* Relationship between clinical quality and patient experience: analysis of data from the english quality and outcomes framework and the National GP Patient Survey. *Ann Fam Med* 2013;**11**:467-72. http://dx.doi.org/10.1370/afm.1514

280. Chang JT, Hays RD, Shekelle PG, MacLean CH, Solomon DH, Reuben DB*, et al.* Patients' global ratings of their health care are not associated with the technical quality of their care. *Ann Intern Med* 2006;**144**:665-72.

281. Roland M, Elliott M, Lyratzopoulos G, Barbiere J, Parker RA, Smith P*, et al.* Reliability of patient responses in pay for performance schemes: analysis of national General

Practitioner Patient Survey data in England. *BMJ* 2009;**339**:b3851. http://dx.doi.org/10.1136/bmj.b3851

282. Department of Health. *National Quality Requirements in the Delivery of Out-of-hours Services*. London, UK: Department of Health; 2006.

283. Tsianakas V, Maben J, Wiseman T, Robert G, Richardson A, Madden P, *et al.* Using patients' experiences to identify priorities for quality improvement in breast cancer care: patient narratives, surveys or both? *BMC Health Serv Res* 2012;**12**:271. http://dx.doi.org/10.1186/1472-6963-12-271

284. Greaves F, Pape UJ, King D, Darzi A, Majeed A, Wachter RM, *et al.* Associations between internet-based patient ratings and conventional surveys of patient experience in the English NHS: an observational study. *BMJ Qual Saf* 2012; 10.1136/bmjqs-2012-000906. http://dx.doi.org/10.1136/bmjqs-2012-000906

285. Greaves F, Millett C, Nuki P. England's Experience incorporating "anecdotal" reports from consumers into their national reporting system: lessons for the United States of what to do or not to do? *Med Care Res Rev* 2014;**71**:65s-80s. http://dx.doi.org/10.1177/1077558714535470

286. Greaves F, Laverty AA, Pape U, Ratneswaren A, Majeed A, Millett C. Performance of new alternative providers of primary care services in England: an observational study. *J R Soc Med* 2015; 10.1177/0141076815583303. http://dx.doi.org/10.1177/0141076815583303

Appendices

[See appended document]