

Does the method of administration influence the UK GMC patient questionnaire ratings?

S.H. Richards¹, J.L. Campbell² and A. Dickens³

¹Senior Lecturer in Primary Care, Primary Care Research Group, Institute of Health Services Research, Peninsula College of Medicine & Dentistry, University of Exeter, Exeter, UK

²Professor of General Practice and Primary Care, Primary Care Research Group, Institute of Health Services Research, Peninsula College of Medicine & Dentistry, University of Exeter, Exeter, UK

³Research Fellow, Primary Care Research Group, Institute of Health Services Research, Peninsula College of Medicine & Dentistry, University of Exeter, Exeter, UK

Aim: To examine whether different methods of administration of a patient questionnaire (PQ) assessing a doctor's professional performance can influence the quality of data collected. **Background:** Patient surveys are important information sources regarding a doctor's professional performance. Previous research across a range of patient outcomes suggests that the method of administration can influence response rates and ratings, although no data that explore how this might influence patients' assessments of a doctor's professional performance are available. **Methods:** A volunteer sample of 13 UK doctors from primary care settings undertook a cross-sectional patient survey. Seven doctors distributed patient surveys using exit and touch-tone telephone administration, and six doctors used exit and postal administration. A consecutive sample of up to 40 patients per doctor completed each method of administration; postal and touch-tone surveys were administered retrospectively, whereas the exit surveys were administered prospectively. The GMC (General Medical Council) PQ included nine performance evaluation questions. We examined the response rates, item completion rates and response profiles of exit survey responses with those obtained from either postal or touch-tone telephone methods. **Results:** The administration methods influence the quality and comparability of data obtained. The exit survey response rate was higher than touch-tone (82.1% versus 37.5%; $P < 0.0001$) or postal survey (no reminder) (76.7% versus 60.8%; $P < 0.0001$) administration, but comparable to the postal survey after one reminder (78.3% versus 75.0%; $P = 0.39$). Item completion rates were comparable for exit and touch-tone surveys (highest rates of missing data 4.1% and 4.6%, respectively), but exit surveys resulted in more missing values than postal surveys (10.3 to 11.7% versus 1.1 to 3.9%). Response profiles varied. Both touch-tone and postal surveys yielded significantly more critical ratings than the exit survey (three of nine items). **Conclusion:** Mixing administration methods requires caution when the data are used to create benchmarks against which a doctor's performance is assessed.

Key words: doctor professional performance; method of administration; patient questionnaire; postal survey; post-consultation exit survey; touchtone telephone survey

Received 13 May 2010; accepted 29 June 2010; first published online 12 August 2010

Correspondence to: S.H. Richards, Senior Lecturer in Primary Care, Peninsula College of Medicine & Dentistry – Primary Care, Smeall Building, St Luke's Campus, Magdalen Road, Exeter EX1 2LU, UK. Email: suzanne.richards@pms.ac.uk

© Cambridge University Press 2010

Introduction

UK government policy will shortly require all doctors to regularly demonstrate their fitness to practice as a requirement of their continued registration (Donaldson, 2006; 2008). Although there is still some uncertainty regarding the optimal methods for gaining evidence on a doctor's professional performance, the current proposals (Donaldson, 2008) include multi-source feedback (Wood *et al.*, 2006; Lockyer and Clyman, 2008). Patients and their families are important sources of information regarding a doctor's performance, and their experiences are most commonly captured through post-consultation 'exit' surveys (Lockyer and Clyman, 2008). However, when used in high-stakes assessment such as revalidation, it is vital that data obtained from patient surveys have robust psychometric properties (Schuwirth *et al.*, 2002).

A recent systematic review of patient survey instruments suitable for use when assessing practising doctors identified six questionnaires that had published some evidence regarding their psychometric characteristics (Evans *et al.*, 2007). This review concluded that there was only limited evidence supporting the psychometric properties of most measures and, in particular, that of construct validity. The authors noted considerable variation in the different methods by which the self-completed patient surveys were administered; some tools were administered as post-consultation 'exit' surveys or as a postal survey. Importantly, no empirical data are available comparing the impact of the method of administration on data quality. The authors concluded that there was potential for the different methods of data collection to result in variations in the evaluations obtained from service users, which, in turn, has the potential to undermine attempts to develop benchmarks of performance. This conclusion is consistent with findings from several reviews of empirical work comparing respondent ratings and the quality of data collected (eg, response rates and item completion rates) across a range of different patient-reported outcome measures (McCull *et al.*, 2001; Bowling, 2005). Such reviews tend to compare studies evaluating different modes of questionnaires, such as self-completion versus interview administration or different methods for completing questionnaires, such as interview-administered questionnaires completed over the telephone or via a face-to-face

interview. When considering self-completion surveys, while postal questionnaires have been widely evaluated, particularly in comparison with interview administration, exit survey methodology or questionnaires completed via interactive voice response (IVR) with automated telephone lines ('touch-tone' telephone) have not been widely investigated or compared (Bowling, 2005).

The UK General Medical Council (GMC) is currently developing a Patient Questionnaire (PQ) that might be suitable for use in the revalidation process. We recently reported on the findings of preliminary testing of this questionnaire in a large, volunteer sample of doctors (Campbell *et al.*, 2008), concluding that it was acceptable and reliable, and had the potential to discriminate a range of performance. This paper addresses the gap in empirical evidence by exploring the effects of the method of administration on data quality in surveys providing information on the professional performance and practice of individual doctors. We examine self-completion methods of administration of the PQ to explore whether or not the data collection process can result in variations in the evaluations made by service users. Assuming that a post-consultation, exit survey is the preferred approach for maximising data collection and ensuring the attributability of the responses to a specific doctor (Campbell *et al.*, 2008), our primary aim was to examine the response rates, item completion rates and response profiles of exit survey responses with that obtained from either postal administration or IVR touch-tone telephone administration of the PQ. A secondary aim was to explore the impact of a reminder questionnaire on the response rate for postal survey administration.

Methods

This study was nested within a large, cross-sectional exit survey undertaken to assess the utility of the GMC PQ (13 754 patients completed a survey after attending one of 380 participant doctors). Detailed reporting of the sampling methods, questionnaire development and findings regarding its psychometric properties are published elsewhere (Campbell *et al.*, 2008). Although NHS Research Ethics Committee guidance was sought prior to the conduct of the survey work, we were advised that a formal submission was not required.

Primary Health Care Research & Development 2011; 12: 68–78

Notwithstanding this, the study was implemented using methods that would have been required by an independent ethical review. Thus, we ensured that doctor and patient participation in the survey work was voluntary, that individuals were fully informed of the purpose of the study, and that all data obtained from individual doctors and patients was anonymised prior to being passed to the research team for analysis.

Study design

A cross-sectional survey was undertaken to address our primary aim, involving two comparisons of the methods for administering the self-completed PQ. The first comparison was that of an exit survey versus touch-tone telephone completion of the PQ, while the second comparison was that of an exit survey versus a postal survey. Data collection for our secondary aim, to explore the impact of a reminder questionnaire on the response rate, was restricted to postal survey administration due to logistical constraints. We selected postal rather than touch-tone methods as we believed, *a priori*, that touch-tone administration was likely to achieve the lowest response rate. Therefore, while a reminder questionnaire might improve the touch-tone response rate, it would remain substantially lower than that of an exit survey.

Settings and participants

We aimed to recruit 20 general practitioners (GPs) from one Primary Care Trust in Devon, United Kingdom. The 96 GPs working within 20 multi-handed practices (ranging from 2 to 11 doctors) in the trust were approached by letter and provided with detailed information about the study aims and methods. Our intention was that 10 doctors would administer the patient survey via exit and postal methods of administration, and a different sample of 10 doctors would administer the survey using both exit and touch-tone telephone methods. Each doctor implemented two methods of administration to minimise any variation in scores being attributed to differences in individual doctors' performance. Thus, any variance observed can be more confidently attributed to the difference in mode of administration, rather than actual differences in performance between doctors. Study participation was voluntary. All doctors were provided with detailed information

regarding the study before agreeing to take part, and were informed that they could withdraw at any point.

Data collection

The PQ was developed to capture those aspects of Good Medical Practice (General Medical Council, 2006) that are amenable to assessment from the perspective of patients or their families. Figure 1 summarises the question stems and fixed response formats (Likert scales or binary categories) for the nine performance evaluation items (questions 3a–g, 4a, b). The PQ also includes items regarding the respondent's sociodemographic characteristics (age group, gender and ethnicity), the perceived importance of their reason for consulting on a 5-point scale (1 = not very important to 5 very important) and whether the consultation was with their usual doctor.

Postal and touch-tone telephone surveys were administered retrospectively. This involved a consecutive sample of the last 40 patients who had consulted with the doctor in the surgery after exclusions. Patients who had consulted with the doctor more than once in the sampling frame were sampled only once. The clinical team was advised that they could exclude patients to whom they felt it was inappropriate to send a survey, but that such circumstances must be exceptional (eg, the patient having experienced a recent bereavement). Minor modifications of the PQ item word stems were required for touch-tone telephone and postal versions of the survey, as the patient would not have seen the doctor on the 'same day'.

Each patient (or their parent/guardian if aged less than 16 years old) was then sent a pack of the materials including a copy of the PQ and a brief information sheet outlining why they had been sent a questionnaire. The information sheet reassured the recipient that their doctor was not being investigated by the GMC, and that their participation in the survey was voluntary.

For the postal survey method, instructions were included on how to return the completed questionnaire direct to an independent survey organisation. After two weeks, non-responders were sent a reminder questionnaire (clearly distinguishable from the original questionnaire).

For the touch-tone telephone method, individuals were provided with a copy of the PQ and

GMC Patient Questionnaire		
<i>Please rate how good your doctor was today at each of the following:</i>		
3a	Being polite	Evaluative*
3b	Making you feel at ease in his / her presence	Evaluative
3c	Listening to you	Evaluative
3d	Assessing your medical condition	Evaluative
3e	Explaining your medical condition	Evaluative
3f	Involving you in decisions about your treatment	Evaluative
3g	Providing or arranging treatment for you	Evaluative
<i>Please decide how strongly you agree or disagree with the following statements:</i>		
4a	I am confident that this doctor will keep information about me confidential	Agreement [†]
4b	I am confident that this doctor is honest and trustworthy	Agreement

* Evaluative response categories (scale number): 'Poor' (1), 'Less than satisfactory' (2), 'Satisfactory' (3), 'Good' (4), 'Very good' (5).

† Agreement response categories (scale number): 'Strongly disagree' (1), 'Disagree' (2), 'Neutral' (3), 'Agree' (4), 'Strongly Agree' (5).

Figure 1 Item stems and response categories for patient questionnaire

instructions on how to call a free-to-call telephone number to access a secure, automated service. On calling the service, an automated voice ran through the survey instructions, followed by each question and response category, before inviting the participant to use their keypad to select the desired response. Patients who failed to respond to the request to complete a touch-tone PQ were not sent a reminder.

Piloting work supporting the main census survey (Campbell *et al.*, 2008) indicated that a reliable assessment of a doctor's performance required at least 30 completed patient questionnaires for each method of administration per doctor, and that around 80% of patients offered an exit survey would accept and complete it. Thus, to conduct the exit survey, doctors were provided with 40 patient questionnaires to be distributed. Administrative staff gave questionnaires to a consecutive sample of patients (excluding any patients with repeat consultations) reporting at reception prior to their consultation. Patients were asked to complete the questionnaire immediately after seeing the doctor and then place their response in a sealed envelope and return it to a collection point within the surgery reception. Doctors were instructed to begin the exit survey immediately after the postal/touch-tone sample had been selected to minimise the potential for any variations in the resultant responses due to time effects.

Data management and analysis

Each doctor was allocated a unique study identification code. Completed surveys were returned to an independent survey organisation. Each questionnaire was inspected and any text that could personally identify an individual was removed (eg, comments referring to names of healthcare staff, or patients and their family members) prior to data entry. Anonymised data were then passed to the research team for analysis.

Patient characteristics, response rates, item completion rates and response profiles of exit survey responses were compared with those obtained from either touch-tone telephone or postal administration. For postal administration, response rates to the first and reminder questionnaires are presented. Differences in proportions were tested using Pearson's χ^2 -test or Fisher's Exact test. Differences between the mean scores (SD) were tested with the appropriate parametric (*t*-test) or non-parametric statistics (Mann-Whitney *U* test) depending on the distribution of data.

Results

Doctor recruitment

Nineteen of the 96 GPs (19.8%) invited provisionally agreed to participate in the study. Six of
Primary Health Care Research & Development 2011; 12: 68–78

these were from one surgery that had recently merged with another, and a practice-level decision was made to withdraw from the study immediately prior to the surveys taking place as the administrative team was busy reconciling processes and procedures, with no capacity to take on additional work. Thirteen doctors (four female) participated in the study (four doctors from one practice, the remainder from different multi-handed practices); seven were allocated to the exit and touch-tone telephone administration, and six were allocated to the exit and postal administration. Data collection took place in May–July 2006. Six of seven doctors who had agreed to take part returned both exit and touch-tone data, with the seventh returning minimal touch-tone data only. Data relating to this doctor was excluded from all subsequent analysis, and thus 12 of 96 doctors approached contributed data.

Comparing exit and touch-tone PQ responses

Data were received from 287 of 480 (59.8%) patients. The response rate for exit surveys (197/240, 82.1%) was more than double that of touch-tone questionnaires (90/240, 37.5%; $\chi^2 = 99.1$, $P < 0.0001$). When comparing the sociodemographic characteristics of patients as recorded in the completed PQs, the proportion of female patients was significantly higher in the touch-tone questionnaires (59/81, 72.8%) compared with the exit survey (88/162, 54.3%; $\chi^2 = 7.7$, $P = 0.005$). However, there was no difference in the ethnic profile (82/87, 94.3% White British in touch-tone questionnaire versus 178/186, 95.7% in exit survey; $\chi^2 = 0.3$, $P = 0.61$) or age profile (<15 years = 10.6%, 15–20 years = 1.2%, 21–40 years = 15.3%, 41–60 years = 28.2%, over 60 years = 44.7% in touch-tone questionnaire versus 2.4%, 2.4%, 22.0%, 31.7%, 41.5%; $\chi^2 = 9.2$, Fisher's exact test $P = 0.07$). The proportion of patients consulting with their usual doctor was similar between touch-tone (22/89, 24.7%) and exit questionnaires (60/193, 31.1%; $\chi^2 = 1.2$, $P = 0.27$) and the mean scores (SD) for the importance of the patient's visit to the doctor were comparable (4.49 (0.83) versus 4.25 (1.10), $t = 1.83$, $P = 0.07$).

Item completion rates were comparable for exit and touch-tone surveys (Table 1), with the highest rates of missing data being 4.4% in both cases. There was some evidence that the profile of responses varied between exit survey and touch-tone surveys

(Table 2), with patient responses to three items of the touch-tone questionnaire being statistically significantly lower (or more critical) than the exit version ('rate the doctor at making you feel at ease in his/her presence', 'rate the doctor on listening to you' and 'rate the doctor on involving you in decisions about your treatment').

Comparing exit and postal PQ responses

Data were received from 368 of 480 (76.7%) patients. The exit survey response rate of 188 of 240 (78.3%) was comparable to the overall postal response rate (180/240, 75.0%; $\chi^2 = 0.75$, $P = 0.39$). However, the reminder enhanced the response rate of the postal questionnaire; 146 of 240 (60.8%) patients returned a completed PQ after the first mail shot and an additional 34 patients completed the reminder questionnaire. The response rate after only one questionnaire was statistically significantly lower than the exit survey response rate (60.8% versus 78.3%; $\chi^2 = 17.36$, $P < 0.0001$).

When comparing the sociodemographic characteristics of patients as recorded in the completed PQs, the proportion of female patients (103/155, 66.5%) in the postal group was comparable with the exit survey group (95/142, 66.9%; $\chi^2 = 0.006$, $P = 0.94$). There was also no difference in ethnicity (168/180, 93.3% White British in the postal group versus 156/161, 96.9%; $\chi^2 = 2.28$, $P = 0.13$) or age (<15 years = 2.5%, 15–20 years = 3.9%, 21–40 years = 20.5%, 41–60 years = 35.3%, over 60 years = 37.8% in postal group versus 7.6%, 6.3%, 26.4%, 32.6%, 27.1%; $\chi^2 = 8.6$, $P = 0.07$). The proportion of patients consulting with their usual doctor was significantly higher in the postal group (79/176, 44.9%) compared with the exit group (52/167, 31.1%; $\chi^2 = 6.9$, $P = 0.009$). The mean scores (SD) for the respondents' assessments of the importance of the patient's visit to the doctor were, however, comparable (4.46 (0.92) versus 4.28 (0.97), $t = 1.8$, $P = 0.07$).

Exit surveys resulted in more missing values for core performance items compared with postal surveys (10.6% to 11.7% versus 1.1% to 3.9%, Table 3). There was also evidence of differential response profiles (Table 4), with postal responses obtaining significantly more critical ratings of the doctors' performance for three of the nine core PQ items ('rate the doctor on assessing your medical condition', 'rate the doctor on explaining

Table 1 Frequency distribution of exit ($n = 197$) and touch tone ($n = 90$) responses

Patient questionnaire item	Mode	Poor (n ; %)	Less than satisfactory (n ; %)	Satisfactory (n ; %)	Good (n ; %)	Very good (n ; %)	Does not apply (n ; %)	Missing (n ; %)
Being polite	Exit	–	–	3 (1.5)	17 (8.6)	173 (87.8)		4 (2.0)
	Touchtone	1 (1.1)	–	1 (1.1)	12 (13.3)	72 (80.0)		4 (4.4)
Making you feel at ease	Exit	–	1 (0.5)	5 (2.5)	17 (8.6)	171 (86.8)		3 (1.5)
	Touchtone	–	–	3 (3.3)	17 (18.9)	66 (73.3)		4 (4.4)
Listening to you	Exit	–	1 (0.5)	4 (2.0)	15 (7.6)	174 (88.3)	1 (0.5)	2 (1.0)
	Touchtone	–	–	4 (4.4)	18 (20.0)	64 (71.1)		4 (4.4)
Assessing your medical condition	Exit	–	–	5 (2.5)	18 (9.1)	168 (85.3)	2 (1.0)	4 (2.0)
	Touchtone	–	–	4 (4.4)	12 (13.3)	69 (76.7)	1 (1.1)	4 (4.4)
Explaining your condition and treatment	Exit	–	1 (0.5)	7 (3.6)	23 (11.7)	159 (80.7)	3 (1.5)	4 (2.0)
	Touchtone	–	1 (1.1)	5 (5.6)	11 (12.2)	64 (71.1)	5 (5.6)	4 (4.4)
Involved in decisions about your treatment	Exit	2 (1.0)	–	9 (4.6)	22 (11.2)	148 (75.1)	9 (4.6)	7 (3.6)
	Touchtone	–	1 (1.1)	5 (5.6)	18 (20.0)	57 (63.3)	5 (5.5)	4 (4.4)
Providing or arranging treatment for you	Exit	1 (0.5)	1 (0.5)	5 (2.5)	16 (8.1)	149 (75.6)	18 (9.1)	7 (3.6)
	Touchtone	1 (1.1)	–	2 (2.2)	9 (10.0)	67 (74.4)	7 (7.8)	4 (4.4)
		Strongly disagree (n ; %)	Disagree (n ; %)	Neutral (n ; %)	Agree (n ; %)	Strongly agree (n ; %)	Does not apply (n ; %)	Missing (n ; %)
Confident about confidentiality	Exit	1 (0.5)	–	4 (2.0)	46 (23.4)	133 (67.5)	7 (3.6)	6 (3.0)
	Touchtone	–	–	1 (1.1)	18 (20.0)	67 (74.4)		4 (4.4)
Confident that doctor is honest/trustworthy	Exit	1 (0.5)	–	5 (2.5)	26 (13.2)	151 (76.6)	6 (3.0)	8 (4.1)
	Touchtone	–	–	–	14 (15.6)	72 (80.0)		4 (4.4)

Table 2 Response profiles of exit and touch tone telephone responses

Patient questionnaire item	Mode	Mean (SD) ^a	Mann–Whitney U, <i>P</i> -value
Being polite	Exit	4.88 (0.37)	7087.5, <i>P</i> = 0.16
	Touchtone	4.79 (0.58)	
Making you feel at ease	Exit	4.85 (0.46)	7418, <i>P</i> < 0.05
	Touchtone	4.73 (0.52)	
Listening to you	Exit	4.87 (0.44)	7085, <i>P</i> < 0.01
	Touchtone	4.70 (0.56)	
Assessing your medical condition	Exit	4.85 (0.42)	7561, <i>P</i> = 0.13
	Touchtone	4.76 (0.53)	
Explaining your condition and treatment	Exit	4.79 (0.52)	7309.5, <i>P</i> = 0.33
	Touchtone	4.70 (0.64)	
Involved in decisions about your treatment	Exit	4.73 (0.66)	6529.5, <i>P</i> < 0.05
	Touchtone	4.62 (0.66)	
Providing or arranging treatment for you	Exit	4.81 (0.57)	6677, <i>P</i> = 0.72
	Touchtone	4.78 (0.61)	
Confident about confidentiality	Exit	4.68 (0.57)	7444.5, <i>P</i> = 0.31
	Touchtone	4.77 (0.45)	
Confident that doctor is honest/trustworthy	Exit	4.78 (0.54)	7732, <i>P</i> = 0.73
	Touchtone	4.84 (0.37)	

^a Owing to the non-parametric distribution of data, although medians (IQR) are the traditional method for presenting such data, here we present means (SD) as they are more informative of the variation within the data set.

your condition and treatment' and 'rate the doctor on providing or arranging treatment for you').

Discussion

The exit survey was the preferred method of administration against which the other methods of administration were compared. We found that the touch-tone telephone administration was prone to bias through substantially higher non-response rates compared with the exit survey. We acknowledge, however, that this response rate might have been improved had we used a questionnaire reminder to non-respondents. Touch-tone questionnaire respondents were also more likely to be female, although there were no differences in the age distribution, ethnicity and the proportion consulting with their usual doctor or the perceived importance of their reason for consulting. There was also some evidence that the telephone version obtained more critical ratings of the doctor's performance for three of nine PQ items when compared with exit survey methods, although this may have been a consequence of the observed response bias.

The postal survey (after incorporating a reminder) might be a potentially suitable alternative to an exit survey. The response rate and sociodemographic

characteristics of patients returning a postal survey (including a reminder questionnaire) were broadly comparable with exit survey respondents, although a higher proportion of the postal group reported consulting with their usual doctor. There was, however, evidence of non-response bias if the postal survey was not accompanied by a reminder questionnaire. The process by which reminders are generated also has implications for the confidentiality of patient responses. In this study, an external survey organisation collated the patient responses and then informed the doctor's administrative team as to which patient questionnaires (identification numbers) required a reminder. While the doctor might be aware of who had completed questionnaires, more importantly, individual patients' ratings remained anonymous. The high response rate achieved after a reminder may not be secured if patients are requested to return completed questionnaires to their doctor's office (ie, negating the need for an external organisation) due to concerns regarding the confidentiality of their feedback. As the generation of reminder questionnaires is a labour-intensive process, the benefits of achieving a comparable response rate need to be balanced against the workload implications of implementing the postal survey.

There was also evidence that the response profiles varied systematically between postal and

Table 3 Frequency distribution of exit (*n* = 188) and postal (*n* = 180) responses

Patient questionnaire item	Mode	Poor (<i>n</i> ; %)	Less than satisfactory (<i>n</i> ; %)	Satisfactory (<i>n</i> ; %)	Good (<i>n</i> ; %)	Very good (<i>n</i> ; %)	Does not apply (<i>n</i> ; %)	Missing (<i>n</i> ; %)
Being polite	Exit	–	2 (1.1)	–	8 (4.3)	157 (83.5)	1 (0.5)	20 (10.6)
	Postal	–	1 (0.6)	3 (1.7)	11 (6.1)	163 (90.6)	–	2 (1.1)
Making you feel at ease	Exit	–	1 (0.5)	1 (0.5)	9 (4.8)	155 (82.4)	1 (0.5)	21 (11.2)
	Postal	–	1 (0.6)	3 (1.7)	13 (7.2)	161 (89.4)	–	2 (1.1)
Listening to you	Exit	–	1 (0.5)	2 (1.1)	13 (6.9)	151 (80.3)	1 (0.5)	20 (10.6)
	Postal	1 (0.6)	1 (0.6)	2 (1.1)	19 (10.6)	152 (84.4)	2 (1.1)	3 (1.7)
Assessing your medical condition	Exit	–	2 (1.1)	3 (1.6)	18 (9.6)	138 (73.4)	5 (2.7)	22 (11.7)
Explaining your condition and treatment	Postal	1 (0.6)	2 (1.1)	7 (3.9)	36 (20.0)	128 (71.1)	4 (2.2)	2 (1.1)
	Exit	–	–	5 (2.7)	14 (7.4)	142 (75.5)	6 (3.2)	21 (11.2)
Involved in decisions about your treatment	Postal	–	4 (2.2)	8 (4.4)	32 (17.8)	126 (70.0)	6 (3.3)	4 (2.2)
	Exit	–	1 (0.5)	4 (2.1)	26 (13.8)	120 (63.8)	15 (8.0)	22 (11.7)
Providing or arranging treatment for you	Postal	3 (1.7)	2 (1.1)	5 (2.8)	36 (20.0)	119 (66.1)	11 (6.1)	4 (2.2)
	Exit	1 (0.5)	–	2 (1.1)	12 (6.4)	129 (68.6)	23 (12.2)	21 (11.2)
	Postal	2 (1.1)	–	4 (2.2)	27 (15.0)	125 (69.4)	15 (8.3)	7 (3.9)
Confident about confidentiality	Exit	3 (1.6)	–	2 (1.1)	30 (16.0)	129 (68.6)	2 (1.1)	22 (11.7)
	Postal	1 (0.6)	1 (0.6)	2 (1.1)	42 (23.3)	128 (71.1)	3 (1.7)	3 (1.7)
Confident that doctor is honest/trustworthy	Exit	3 (1.6)	–	2 (1.1)	28 (14.9)	134 (71.3)	–	21 (11.2)
	Postal	2 (1.1)	–	–	40 (22.2)	134 (74.4)	2 (1.1)	2 (1.1)

Table 4 Response profiles of exit and postal responses

Patient questionnaire item	Mode	Mean (SD) ^a	Mann–Whitney <i>U</i> , <i>P</i> -value
Being polite	Exit	4.92 (0.39)	14 498.5, <i>P</i> = 0.38
	Postal	4.89 (0.41)	
Making you feel at ease	Exit	4.92 (0.36)	14 338, <i>P</i> = 0.32
	Postal	4.88 (0.42)	
Listening to you	Exit	4.88 (0.41)	14 093, <i>P</i> = 0.30
	Postal	4.83 (0.52)	
Assessing your medical condition	Exit	4.81 (0.52)	12 306.5, <i>P</i> < 0.01
	Postal	4.66 (0.67)	
Explaining your condition and treatment	Exit	4.85 (0.44)	11 744, <i>P</i> < 0.005
	Postal	4.65 (0.68)	
Involved in decisions about your treatment	Exit	4.75 (0.53)	11 493, <i>P</i> = 0.11
	Postal	4.61 (0.77)	
Providing or arranging treatment for you	Exit	4.86 (0.48)	10 189.5, <i>P</i> < 0.05
	Postal	4.73 (0.64)	
Confident about confidentiality	Exit	4.72 (0.67)	13 588.5, <i>P</i> = 0.31
	Postal	4.70 (0.58)	
Confident that doctor is honest/trustworthy	Exit	4.74 (0.66)	14 163, <i>P</i> = 0.42
	Postal	4.73 (0.58)	

^a Owing to the non-parametric distribution of data, although medians (IQR) are the traditional method for presenting such data, here we present means (SD) as they are more informative of the variation within the data set.

exit surveys, with postal responses being more critical of the doctor's performance for three of the nine core items. This disparity may be a result of recall bias, as the time elapsed between the consultations and completion of questionnaires varied. Patients were invited to complete a postal questionnaire 2–14 days after their consultation, rather than immediately afterwards when completing an exit survey. Thus, the postal survey allowed patients more time to reflect on the consultation, and possibly to experience new or additional health problems that may alter their judgement of the doctor. An alternative explanation is that an individual completing a postal survey in the privacy of their own home may feel less inhibited towards providing critical assessments than those completing questionnaires within a surgery waiting room. Although our data cannot identify the precise cause of the disparity between ratings, our findings do suggest that the postal and exit surveys may be measuring slightly different concepts, and that a cautious approach to mixing data from different methods of administration should be adopted.

By comparing data obtained from doctors practising in similar settings (primary care) using two different methods of administration (eg, exit versus postal, or exit versus touch-tone), it is

unlikely that the differences observed between the methods are solely attributable to differences in performance between doctors or in the settings in which they practise. Given that different patients contributed to the two methods of administration tested per doctor, we cannot rule out that individual doctors' performance may have varied between the methods of administration, which, in turn, resulted in the different response profiles. Neither can we conclusively rule out patient selection bias between doctors, which is a limitation of this work. Although doctors were actively encouraged not to exclude patients from the sampling frames unless there were exceptional circumstances, we did not collect sufficient data to explore this fully. In considering the potential impact of case mix, however, no differences were observed between groups in respect of respondents' perceptions of the importance of their visit to the doctor. Respondents completing a postal survey were, however, more likely to report that their consultation had been with their usual doctor. Notwithstanding this, our finding that both postal and touch-tone administration resulted in more critical assessments of a doctor's performance compared with exit survey methodology provides the first empirical support for the conclusions

Primary Health Care Research & Development 2011; **12**: 68–78

drawn (in the absence of comparative data) in a recent review of six self-completion patient survey instruments assessing a doctor's performance (Evans *et al.*, 2007).

A final consideration, potentially of some importance, relates to the issue of attributability. Given the context in which patients' views are being sought and thus the significance of the judgements being made, it is important that these judgements relate to the performance of the specific doctor who is the target of the assessment. An exit survey, specifically focussing on the performance of the doctor just seen is more likely to address the issue of attributability than the two alternative methods described here, both of which are temporally more distant and less directly associated with a specific doctor's performance than the post-consultation exit survey. It is plausible that patients receiving postal or touch-tone surveys may have consulted with another doctor in the time elapsed between the index consultation and being in receipt of a questionnaire. This scenario allows for the patient to inadvertently complete the assessment for the 'wrong' doctor.

Further research is needed to explore the effects of the method of administration on data quality in surveys providing information on the professional performance and practice of individual doctors. To improve the generalisability of our research findings, this study should be replicated across a range of clinical settings including secondary care specialties. It is also vital that different patient groups contribute, as our sample was relatively homogenous, and, in particular, lacked ethnic and socio-economic diversity. Such research should also focus on why more critical ratings were achieved for postal and touch-tone telephone surveys compared with exit survey methodology, and seek to examine the impact of new methods of data collection, such as computer technology. There is increasing interest in the use of mixed modes of data collection in survey design, so that data collection procedures can be more flexible, and tailored to the context in which the surveys are administered to maximise response rates and improve the speed of data collection (Groves *et al.*, 2004). An improved understanding of why methods can elicit different responses and how such differences can be accommodated in the analysis of survey data is

essential to guide our understanding of how benchmarks might be generated and used to support standard setting. At present, caution is required towards mixing administration methods when creating and applying benchmarks against which a doctor's performance is assessed until further research is undertaken. This is of particular importance should such patient questionnaires become a routine component of multi-source feedback in the UK's current plans for the revalidation of doctors.

Acknowledgements

This study was funded by the United Kingdom General Medical Council. We gratefully acknowledge the general practitioners who supported this study, and the patients who completed questionnaires. We also thank CFEP-UK surveys, Exeter, who developed the touch-tone telephone service, and supported the data handling and processing for this study.

References

- Bowling, A.** 2005: Mode of questionnaire administration can have serious effects on data quality. *Journal of Public Health* 27, 281–91.
- Campbell, J., Richards, S., Dickens, A., Greco, M., Narayanan, A. and Brearley, S.** 2008: Assessing the professional performance of UK doctors: an evaluation of the utility of the General Medical Council patient and colleague questionnaires. *Quality and Safety in Health Care* 17, 187–93.
- Donaldson, L.** 2006: *Good doctors, safer patients: proposals to strengthen the system to assure and improve the performance of doctors and to protect the safety of patients: a report*. London: Chief Medical Officer, Department of Health.
- Donaldson, L.** 2008: *Medical revalidation – principles and next steps*. Leeds, England: Chief Medical Officer, Department of Health (Great Britain).
- Evans, R., Edwards, A., Evans, S., Elwyn, B. and Elwyn, G.** 2007: Assessing the practising physician using patient surveys: a systematic review of instruments and feedback methods. *Family Practice* 24, 117–27.
- General Medical Council.** 2006: *Good medical practice*. London: General Medical Council.
- Groves, R., Fowler, F., Couper, M., Lerchen, M., Singer, E. and Tourangeau, R.** 2004: *Survey methodology*. New Jersey, Canada: John Wiley & Sons, 137–68.

- Lockyer, J.** and **Clyman, S.** 2008: Multi source feedback (360 degree evaluation). In Holmboe, E. and Hawkins, R., editors, *Practical guide to the evaluation of clinical competence*, first edition. Philadelphia: Mosby Elsevier, 75–85.
- McColl, E., Jacoby, A., Thomas, L., Soutter, J., Bamford, C., Steen, N., Thomas, R., Harvey, E., Garratt, A. and Bond, J.** 2001: Design and use of questionnaires: a review of best practice applicable to surveys of health staff and patients. *Health Technology Assessment* 5(31).
- Schuwirth, L., Southgate, L., Page, G., Paget, N., Lescop, J., Lew, S., Wade, W. and Baron-Maldonado, M.** 2002: When enough is enough: a conceptual basis for fair and defensible practice performance assessment. *Medical Education* 36, 925–30.
- Wood, L., Hassell, A., Whitehouse, A., Bullock, A. and Wall, D.** 2006: A literature review of multi-source feedback systems within and without health services, leading to 10 tips for their successful design. *Medical Teacher* 28, e185–91.