BMJ Open

The Norwegian patient experiences with GP questionnaire (PEQ-GP): reliability and construct validity following a national survey

Olaf Holmboe, Hilde Hestad Iversen, Kirsten Danielsen, Oyvind Bjertnaes

To cite: Holmboe O, Iversen HH, Danielsen K, *et al*. The Norwegian patient experiences with GP questionnaire (PEQ-GP): reliability and construct validity following a national survey. *BMJ Open* 2017;**7**:e016644. doi:10.1136/bmjopen-2017-016644

► Prepublication history and additional material for this paper are available online. To view please visit the journal (http://dx.doi.org/10.1136/bmjopen-2017-016644).

Received 3 March 2017
Revised 1 June 2017
Accepted 28 July 2017

Division of Health Services, Norwegian Institute of Public Health, Oslo, Norway

Correspondence to
Olaf Holmboe;
olaf.holmboe@fhi.no

ABSTRACT

Objectives The aim of this study was to test the reliability and validity of a new questionnaire for measuring patient experiences with general practitioners (PEQ-GP) following a national survey.

Setting Postal survey among patients on any of 500 GPs patient lists in Norway. GPs were stratified by practice size and geographical criteria.

Participants 4964 patients who had at least one consultation with their regular GP in the foregoing 12 months were included in the study. The patients were randomly selected after the selection of GPs. 2377 patients (49%) responded to the survey.

Primary and secondary outcome measures The items were assessed for missing data and ceiling effects. Factor structure was assessed using exploratory factor analyses. Reliability was tested with item–total correlation, Cronbach's alpha and test–retest correlations. Item discriminant validity was tested by correlating items with all scales. Construct validity was assessed through associations of scale scores with health status, the patients' general satisfaction with the services, whether the patient had been incorrectly treated by the GP and whether the patient would recommend the GP to others.

Results Item missing varied from 1.0% to 3.1%, while ceiling effects varied from 16.1% to 45.9%. The factor analyses identified three factors. Reliability statistics for scales based on these three factors, and two theoretically derived scales, showed item–total correlations ranging from 0.63 to 0.85 and Cronbach's alpha values from 0.77 to 0.93. Test–retest correlation for the five scales varied from 0.72 to 0.88. All scales had the expected association with other variables.

Conclusions The PEQ-GP has good evidence for data quality, internal consistency and construct validity. The PEQ-GP is recommended for use in local, regional and national surveys in Norway, but further studies are needed to assess the instrument's ability to detect differences over time and between different GPs.

Strengths and limitations of this study

- The PEQ-GP was developed and validated according to the standard scientific procedures of the national patient-reported experience programme in Norway
- Tests of the validity and reliability of the PEQ-GP was conducted in a large national survey in Norway
- The lack of information about non-respondents precluded the possibility of assessing non-response bias
- The ability of the PEQ-GP to identify important differences between providers or over time was not assessed

BACKGROUND

Patients are increasingly being involved in the planning, implementation and evaluation of healthcare services. The latter involves measurements and instruments related to patient-reported quality, including patient-reported experiences, patient-reported outcomes and patient-reported safety. The importance of patient-reported experiences is acknowledged by international organisations like the Organisation for Economic Co-Operation and Development (OECD) and WHO,[1 2] and supported by research showing a positive correlation between patient experiences and clinical outcomes and patient safety.[3 4] Numerous instruments for the measurement of patient-reported experiences have been validated, and standardised measurements are increasingly being used for high-stake purposes like external quality indicators and pay for performance.

In Norway, the Norwegian Institute of Public Health is responsible for national surveys of patient-reported experiences with healthcare services. The purpose of the surveys is systematic measurement of patient experiences, as basis for accountability, hospital management, quality improvement and patients' choice of healthcare provider. Hence, the surveys are of interest to both patients and decision makers on all levels, including the national quality indicator system. Following Donabedians perspective on quality[5] patient experiences surveys for a great part inform the structural and processual aspects of quality, while

outcome measures have been less focused. The institute has developed, validated and published a large number of questionnaires for a wide range of patient groups and healthcare services.[6–11] However, most instruments have been developed and validated for patient groups in specialist healthcare, and so far, only one instrument has been developed for primary healthcare services.[8] General practitioners (GPs) are a cornerstone of the Norwegian healthcare system, with almost all inhabitants having a regular GP as part of the primary healthcare system. GPs diagnose and treat patients for a number of conditions, refer patients who need it to specialist healthcare and are expected to coordinate and cooperate with other services both in primary and specialist healthcare. Consequently, patient experiences with GPs are of major interest, and a literature review was conducted to identify potential instruments.[12] This review identified several relevant instruments, two of which were validated in Norway, the Patient Experience Questionnaire and the EUROPEP.[13 14] Both questionnaires were different from the standard format of Norwegian national patient experiences questionnaires, but the topics and questions were included in the development project. The PEQ instrument is narrow in its scope and is designed for evaluating a specific consultation which is not a feasible approach in a national survey. Results from a survey using EUROPEP in Norway indicates that this instrument yielded high proportions of item non-response, large ceiling effects and low GP-level reliability for several items[15] and similar criticism was raised in a Danish study.[16] In addition, the EUROPEP instrument does not include questions about patient safety nor coordinated care across providers, which are important issues in the Norwegian healthcare debate. On the background of these findings, it was decided to establish a development project based on the standard development and validation methodology used in the national patient experience programme.[6–11] Activities in the development project included the review, meetings in a reference group of health personnel and other relevant stakeholders, cognitive interviews with patients and a pilot survey,[17] producing a test version of the PEQ-GP for inclusion in a national validation survey in 2014.

The aim of this study was to test the reliability and construct validity of the PEQ-GP based on the national survey comprising 4964 patients across Norway.

METHODS

Setting

All residents in Norway are entitled to a regular GP. Each municipality establishes agreements with GPs to serve their population. The GPs have a list of patients for whom they have medical responsibility. Available data indicate that more than 99% of the Norwegian population are on a regular GP's patient list.[18] Norwegian GPs are gate-keepers for the national insurance scheme, and patients are referred from a GP to specialised medical care when needed. Norwegian GPs' practices are, in general, small units, employing on average approximately three GPs. Normally, there are also one or more receptionists as well as staff for sampling and analysing simple tests at the GP practice. Other healthcare workers like nurses or physiotherapists are rare in Norwegian GP offices.

Development of the questionnaire

To identify important topics, we assessed reviews of the literature[12 19] and conducted a thorough assessment of the content of several questionnaires[13 14 20–22] as well as questionnaires used in national programme in Sweden and the USA. Consultations with a reference group comprising GPs, researchers and representatives from health authorities and patient organisations also added to the content of the questionnaire. Several drafts of the questionnaire were discussed with this group.

The questionnaire was tested through cognitive interviews with patients at different stages in the process. Cognitive interviewing is a standard procedure in the development process of questionnaires. The purpose is to find out how the questionnaire functions cognitively in the target group, for instance, how patients interpret items and the adequacy of response categories. First, we conducted eight face-to-face interviews and nine telephone interviews. The questionnaire was revised on the basis of these interviews and discussions with the reference group. The revision was extensive, thus we conducted further face-to-face interviews with 11 patients based on the new draft. The second round showed that the questionnaire functioned well, with some remaining issues to solve. Some patients found it difficult to know who to evaluate when their regular GP was on long-term leave from the practice. Furthermore, questions about cooperation were difficult to answer because only a few patients perceived any need for coordinated care. This led to minor revisions of items and leading texts in the questionnaire. Before the validation survey, the revised version was tested in a pilot study. The pilot sample was 150 patients from each of five randomly selected GPs' lists. The results of the pilot study were discussed with the expert group, and the questionnaire was revised before the validation survey in 2014.

The PEQ-GP used in the national validation study contained 30 items, covering topics like accessibility, the GP's medical skills and relationship with the patient, organisation of the GP's office, coordination of care, general satisfaction, patient enablement and incorrect treatment. Most of the items had a 5-point scale ranging from 'Not at all' to 'To a large extent'. The questionnaire also included background variables like the presence of a long-term medical condition, self-reported health status and whether the patient had received help answering the questionnaire. Finally, we included an open-ended item for comments about the survey (see online supplementary file, English version of the questionnaire).

Table 1 Sample descriptives

	All respondents		Retest sample	
	Frequency	%	Frequency	%
Gender				
Male	558	23.6	40	22.2
Female	1809	76.4	140	77.8
Age				
0–5 years	106	4.5	4	2.2
6–15 years	153	6.5	8	4.4
16–19 years	66	2.8	1	0.6
20–29 years	143	6.0	3	1.7
30–49 years	565	23.9	28	15.6
50–66 years	701	29.6	64	35.6
67–79 years	476	20.1	57	31.7
80–89 years	133	5.6	14	7.8
90 years or more	24	1.0	1	0.6
Education				
None	11	0.5	0	0.0
Primary school	547	26.2	32	19.0
High school	914	43.8	79	47.0
University graduate	477	22.9	39	23.2
University postgraduate	137	6.6	18	10.7
Born in Norway				
No	225	9.6	10	5.6
Yes	2126	90.4	170	94.4
Self-reported health status				
Excellent	250	10.6	19	10.6
Very good	668	28.3	48	26.7
Good	801	34.0	69	38.3
Fairly good	498	21.1	36	20.0
Poor	142	6.0	8	4.4
Long-standing health problem (>6 months)				
None	749	32.0	54	30.0
Yes, one	817	34.9	49	27.2
Yes, two	468	20.0	24	13.3
Yes, three or more	304	13.0	53	29.4
Number of consultations in last 12 months				
1–2	751	31.6	42	23.3
3–4	493	20.7	32	17.8
five or more	1133	47.7	106	58.9
Years the patient has been on this GP's patient list				
0–5 years	1058	44.7	65	36.1
6–12 years	637	26.9	48	26.7
13 years or more	672	28.4	67	37.2
Who filled in the questionnaire?				
Patient	2077	88.2	167	92.8
Next of kin to children<16 years	242	10.2	12	6.7
Next of kin to others	35	1.5	1	0.6

Sample

Only regular GPs and their patients were included in the study. Private primary care physicians and patients not on a regular GP list were excluded. The sampling plan aimed to give a nationally representative sample, and had a three-stage design. First, practices were stratified by number of GPs at the practice and municipality type and randomly selected. Second, we selected up to four GPs from each of the selected practices: if the practice had five or more GPs, four of them were randomly selected; while in smaller practices, all GPs were included. Third, we selected randomly 10 patients from each of the GPs' lists. All of the patients had to have at least one consultation with their GP between May 2013 and May 2014.

Data collection

A total of 4964 persons were mailed a questionnaire with a cover letter describing the purpose of the survey, a prepaid return envelope and an option to answer electronically. To ensure that the patients evaluated the intended GP, the cover letter included the name of the patient's regular GP. For persons below 16 years (age of consent) and other persons who had difficulties filling in the questionnaire, the caregivers/next of kin were asked to respond on their behalf. A reminder including a new questionnaire was sent to those who had not responded within 3 weeks.

Background information about the respondents was collected from public registries. These data included the selected patients' age, gender, how long the patient had been on the GP's patient list, number of consultations in the last 12 months, diagnoses in the last 12 months, level of education and country of birth.

Statistical analyses

Items were assessed for item-missing and ceiling effects, the latter defined as the percentage of patients selecting the most positive response to each item. Patient experience items with less than 20% missing values (item missing and not applicable) were entered into an exploratory factor analysis with promax rotation to assess the underlying structure of the questionnaire.[23] Two additional scales were constructed based on theoretical assumptions. The first consisted of three items inspired by the Patient Enablement Instrument.[24] These were considered outcome measures and not patient experiences, and were therefore excluded from the initial factor analysis. Two items evaluating coordination of care and cooperation with other health services were excluded from the factor analyses because they were relevant only to some of the participants and therefore had high levels of non-valid responses. However, they were included as a separate scale because this is a core topic in the Norwegian health policy. Three items related to accessibility were excluded from psychometric analysis because they were conceptually distinct and with different response categories than all other items (for instance, waiting time in days/weeks). These are not reported here.

Table 2 Descriptive statistics for scales and items, internal consistency and test–retest reliability

Item no	Item	Missing n(%)	Mean	SD	% Ceiling	Cronbach's alpha/ item total correlation	Test–retest correlation
	GP	24 (1.0)	77.4	17.64		0.932	0.88
7	GP takes you seriously	23 (1.0)	4.3	0.84	45.4	0.809	
8	GP has enough time to you	27 (1.1)	4.0	0.93	31.4	0. 752	
9	GP speaks so that you can understand her/him	25 (1.1)	4.3	0.75	45.9	0. 741	
10	GP is competent	35 (1.5)	4.2	0.74	36.3	0. 765	
11	GP is interested in your situation	36 (1.5)	4.1	0.85	35.6	0. 846	
12	GP involves you as much as you want in decisions about you	44 (1.9)	4.1	0.82	32.8	0. 798	
13	GP gives sufficient information about health problems and treatment	215 (9.0)	4.1	0.88	36.0	0.803	
14	GP provides sufficient information about use and side effects of medication	383 (16.1)	3.6	1.04	21.0	0. 652	
	Auxiliary staff	23 (1.0)	76.3	17.83		0.845	0.80
15	Practice well organised	34 (1.4)	3.9	0.83	20.4	0.639	
16	Auxiliary staff forthcoming and competent	30 (1.3)	4.0	0.80	28.6	0.787	
17	Met with politeness and respect at reception	70 (2.9)	4.3	0.81	43.5	0.714	
	Accessibility	116 (4.9)	64.8	27.75		0.772	0.72
4	Acceptability of waiting time for acute consultations	413 (17.4)	3.8	1.24	38.8	0.629	
6	Acceptability of waiting time for normal consultations	260 (10.9)	3.4	1.18	18.8	0.629	
	Enablement	52 (2.2)	65.9	21.57		0.889	0.74
23	Contact with GP helps you understand your illness	56 (2.4)	3.7	0.92	19.1	0.832	
24	Contact with GP helps you cope with your illness	64 (2.7)	3.7	0.94	17.5	0.788	
25	Contact with GP helps you keep yourself healthy	73 (3.1)	3.5	1.00	16.1	0.734	
	Cooperation	396 (16.7)	74.4	21.18		0.870	0.80
20	GP coordinates health services well	534 (22.5)	4.0	0.88	28.1	0.771	
21	GP cooperates well with other services	632 (26.6)	4.0	0.92	30.5	0.771	
	Items excluded after initial factor analysis						
1	Phone accessibility	52 (2.2)	3.8	1.09	33.0		
18	Waiting time in waiting room beyond appointment time	26 (1.1)	3.1	1.03	5.5		
19	Privacy ensured in conversations	74 (3.1)	4.0	0.97	32.7		

Items are scored 1 to 5 and scales are scored 0 to 100 where higher scores represent better experiences. To achieve a scale score, the responder had to give a valid answer to at least half of the items pertaining to the scale. GP, general practitioner.

Internal consistency was assessed by item–total correlation and Cronbach's alpha. The former measures the strength of association between an item and its scale, and levels above 0.4 are considered acceptable.[25] The latter assesses the overall correlation between items within a scale. For a scale to be considered reliable, an alpha value of 0.7 is considered acceptable.[25] Two hundred and seventy consenting responders were mailed a second questionnaire, enabling us to assess test–retest reliability.

Construct validity was assessed through associations of scale scores with variables known to correlate with patient-reported experiences: self-reported health and patients' overall satisfaction with the services.[6–11] We hypothesised that higher scores on the scales would be associated with higher levels of general satisfaction and patient safety, and with better health, and tested these associations by bivariate correlations. Item discriminant validity was tested by correlating items with all scales, and we expected each item to have a significantly higher correlation with its hypothesised scale than with scales measuring other concepts.[26]

RESULTS

One hundred and seven persons were excluded because they withdrew from the survey, had an unknown address or had died by the time of the survey. We received 2377 responses to the survey (49%), of which 439 were electronic responses. Mean age of the patients was 40 years and 76% were women (table 1). Persons who answered electronically were younger, had higher education and rated their own health as better than patients filling in the paper form (results not shown). Twelve per cent of the responses were made by caregivers or next of kin. Most

Table 3 Factor analysis with loadings (n=1508)

Factors/items	Factor 1	Factor 2	Factor 3
GP			
GP is interested in your situation	0.910	0.011	−0.050
GP takes you seriously	0.862	−0.048	0.030
GP gives sufficient information about health problems and treatment	0.840	0.020	−0.020
GP involves you as much as you want in decisions about you	0.833	0.000	0.010
GP is competent	0.805	0.032	−0.033
GP speaks so that you can understand her/him	0.773	0.027	−0.008
GP has enough time to you	0.745	−0.032	0.095
GP provides sufficient information about use and side effects of medication	0.692	−0.014	−0.021
Auxiliary staff			
Auxiliary staff forthcoming and competent	−0.048	0.984	−0.056
Met with courtesy and respect at reception	−0.032	0.817	0.004
Practice well organised	0.168	0.560	0.128
Accessibility			
Acceptability of waiting time for normal consultations	−0.032	0.000	0.867
Acceptability of waiting time for acute consultations	0.021	0.008	0.722

items had a low level of item-missing and ceiling effects ranged from 6 to 46 (table 2).

The initial factor analysis showed that three items (1, 18 and 19) had factor loadings less than 0.4. These were removed and the final factor analysis produced a three factors solution, explaining 66% of the variation in the included variables (table 3). The first factor consisted of eight items about the GP's medical and relational skills. The second factor consisted of three items about the organisation of the practice and assessments of auxiliary staff, and the third factor was two items assessing the acceptability of waiting time for acute and non-acute consultations, respectively. Reliability statistics for these three scales, together with the Enablement and the Cooperation scales, showed acceptable Cronbach's alphas and item–total correlations: the former ranged from 0.77 for accessibility to 0.93 for the GP scale (table 2). Test–retest correlations for the five scales were high, ranging from 0.72 (accessibility) to 0.88 (GP).

All items had a stronger correlation with their own scale than with any of the other scales (table 4). All correlations were significant ($p<0.001$). High correlations were also observed between the GP, Cooperation and Enablement scales and the items pertaining to the other of these scales (table 4).

The scales were expected to correlate with general satisfaction with GP, patient safety, recommendation of the GP and with the patients' health status. The item about general satisfaction with GP had moderate correlations (0.388–0.438) with the Accessibility and Auxiliary staff scales and high correlations (0.693–0.826) with the rest of the scales (table 5). The same pattern applies to recommendation of the GP. Self-reported health status had low but significant correlations with all five scales (0.041–0.212). Overall, all correlations were significant and in the expected direction.

DISCUSSION

The development of the PEQ-GP followed a standard procedure including a literature review of existing questionnaires, input from an expert panel, cognitive interviews with patients and a pilot study. The national validation study identified five scales with excellent psychometric properties, covering important aspects of the GP service relating to accessibility, evaluation of the GP and auxiliary staff, cooperation between the GP and other services and patient enablement.

Patient-reported experiences usually relate to structures and processes of healthcare, but the PEQ-GP also offers an intermediate outcome indicator through the Enablement scale. Thus, the PEQ-GP includes all aspects of Donabedian's classical structure-process-outcome framework, offering a broad measurement approach of GP services.[5] We are not aware of other patient-reported experience instruments with GPs with the same breadth in scale content. Together with single items not included in the scales, some of which are important in the political discourse, the PEQ-GP covers the most important topics for patients and decision makers on all levels in Norway.

High correlations were found between the GP, Cooperation and Enablement scales. This may be because they all evaluate the work of a particular GP, and thus are closely related. Cooperation was left out of the factor analysis because of high missing values, and Enablement was left out because the items were considered outcome and not experience measures. Although they correlate with the GP scale, they both are conceptually different from it, because they do not assess the relational and patient-oriented work of the GP. Furthermore, patient enablement is a major goal in most western healthcare systems, and thus a useful supplement to the patient experience items. A recent study from England showed the importance of

Table 4 Correlations between items and scales

	GP	Auxiliary staff	Cooperation	Enablement	Accessibility
GP takes you seriously	0.853	0.349	0.654	0.626	0.362
GP has enough time for you	0.818	0.373	0.626	0.584	0.389
GP speaks so that you can understand her/him	0.794	0.366	0.595	0.554	0.309
GP is competent	0.815	0.380	0.657	0.636	0.324
GP is interested in your situation	0.884	0.393	0.703	0.685	0.349
GP involves you as much as you want in decisions about you	0.847	0.387	0.666	0.628	0.369
GP gives sufficient information about health problems and treatment	0.852	0.396	0.679	0.698	0.347
GP provides sufficient information about use and side effects of medication	0.749	0.299	0.564	0.626	0.271
Practice well organised	0.460	0.841	0.477	0.423	0.402
Auxiliary staff forthcoming and competent	0.370	0.908	0.380	0.350	0.334
Met with politeness and respect at reception	0.344	0.876	0.361	0.329	0.326
GP coordinates health services well	0.747	0.471	0.945	0.663	0.368
GP cooperates well with other services	0.735	0.413	0.948	0.645	0.346
Contact with GP helps you understand your illness	0.736	0.385	0.653	0.904	0.342
Contact with GP helps you cope with your illness	0.717	0.398	0.653	0.928	0.354
Contact with GP helps you keep yourself healthy	0.606	0.360	0.547	0.887	0.318
Acceptability of waiting time for acute consultations	0.369	0.358	0.325	0.313	0.911
Acceptability of this waiting time for normal consultations	0.395	0.398	0.373	0.375	0.909

empowerment for patients in a general practice setting.[27] However, more research is needed to evaluate the predictive validity of the Enablement scale, particularly the association between the scale and other outcomes like compliance and health outcome.

The focus on cooperation in Norwegian healthcare is strong, with more resources invested in projects and initiatives to improve cooperation and integration of care. Internationally, cooperation and integration of care are also important topics and specific instruments have been developed and validated.[28–32] Thus, the Cooperation scale should be a part of the PEQ-GP in Norway. However, the Cooperation scale presents two challenges: (1) the items constituting the scale were relevant only to about 75% of the patients; (2) the national scale score was surprisingly high, given the fact that poor cooperation was the diagnosis before the implementation of the national Cooperation Reform. Regarding the first challenge, one approach would be to include this scale only when a particular interest lies in the cooperation topic, which was the reason for including cooperation in the first place. A statistical recommendation is to increase the sample size by 25%

Table 5 Association between scales and other variables. Mean scale scores by response category.

		GP	Auxiliary staff	Cooperation	Enablement	Accessibility
Generally satisfied with GP	Pearson correlation	0.826	0.438	0.707	0.693	0.388
	Sig. (two-tailed)	<0.001	<0.001	<0.001	<0.001	<0.001
	N	2344	2339	1976	2319	2242
Self-reported health status	Pearson correlation	−0.212	−0.041	−0.139	−0.193	−0.142
	Sig. (two-tailed)	<0.001	0.048	<0.001	<0.001	<0.001
	N	2347	2347	1977	2322	2251
Incorrectly treated by GP	Pearson correlation	−0.0465	−0.225	−0.401	−0.375	−0.160
	Sig. (two-tailed)	<0.001	<0.001	<0.001	<0.001	<0.001
	N	2326	2319	1959	2304	2224
Recommend GP to friends or family	Pearson correlation	0.791	0.389	0.661	0.675	0.367
	Sig. (two-tailed)	<0.001	<0.001	<0.001	<0.001	<0.001
	N	2186	2183	1864	2168	2087

at the GP or practice level to compensate for the high levels of item missing. The second problem is probably related to the fact that the cooperation items are formulated rather generally and are substantially quite closely related to the GP items, the latter documented by the large correlations with the GP scale. Future development and research work is needed to test more specific questions about cooperation, with formulations more focused on the joint responsibility for cooperation between GPs and other services.

The PEQ-GP can be used to inform individual doctors about their patients' view of their services, providing a basis for quality improvement for practices and individual GPs. Follow-up studies indicate that feedback through patient evaluations may be used for quality improvement in healthcare.[33–35] Low missing values indicate that the items are acceptable to the patients, and moderate levels of ceiling effects indicate that the instrument could be useful for detecting changes over time and differences between GPs. However, repeated measurements are required to detect changes; thus we are not able to test the sensitivity of the instrument. Another limitation is the low number of respondents for each GP. This makes it difficult to test the appropriateness of using the PEQ-GP scales as a basis for external quality indicators, including the ability to discriminate between GPs and/or practices. Results from the pilot study indicate that there are substantial differences between GPs, but further studies are needed to assess the PEQ-GP's benchmarking abilities at both the GP and practice levels.

The total number of respondents is high. This is a strength of the study, but the response rate of 49% is not optimal. Earlier studies indicate small differences between non-responders and responders. Furthermore, response rate is in itself a poor indicator of non-response bias.[36–38] The lack of background information about non-respondents in the present study is obviously a weakness, and means that we are not able to rule out the possibility of non-response bias. One possible indicator of bias is the high proportion of women among the responders (76%). As we do not have any information about the non-responders, we cannot relate this result to response propensity. Other studies of patient experiences in primary care have reported that around 2/3 of the responders were women.[39 40] In a Danish study, Heje *et al* stated that this proportion reflects the consultation pattern between the genders.[41] In Norway, there is also evidence that a higher proportion of women than men see their GP during any 1 year, and that they have more consultations than men.[42] However, there is no clear evidence in the GP field that women evaluate health services differently than men. We therefore assume that the high proportion of women does not cause any substantial problems for the assessment of the properties of the PEQ-GP.

The retest population differs slightly from the rest of the responders on several variables, implying that test–retest results should not automatically be generalised to the total respondent sample. The purpose of the retest was to measure the reproducibility of the results within the same individuals, as a test of the questionnaires reliability. The results supported the test–retest reliability of the PEQ-GP in this sample, but further studies should try to replicate findings to achieve a more robust knowledge base.

Another weakness of this study is that some patients may have had little contact with their GP. The reasons may be numerous, such as no need for medical services, contact with other GPs or staff changes in the GP office. Limited contact with their GP may affect the patients' ability to evaluate the GP, and furthermore, a long period of time since visiting the GP may cause recall bias or a blurred impression. To reduce this weakness, only patients with at least one consultation in the last 12 months were included in the study.

CONCLUSION

The PEQ-GP includes important aspects of patient experiences with GPs. The questionnaire has evidence for data quality, reliability and construct validity. The PEQ-GP is recommended for future studies designed to assess patient experiences with GPs in Norway, but further research is needed to test the appropriateness of using the PEQ-GP as a basis for external quality indicators and in other countries.

Acknowledgements The authors thank Tomislav Dimoski for managing the data collection, developing the software used in data collection and management (The FS system) and carrying out the technical aspects of the data collection, and Marit Skarpaas and Inger Opedal Paulsrud for administrative help in data collection. We also thank Jon Helgeland for calculating the sample model and Andrew Garratt for advice in the translation process.

Contributors All authors participated in the planning process. OH performed the statistical analyses and drafted the manuscript. HHI, OAB and KD critically revised the manuscript draft and approved the final version of the manuscript. OH was the project manager for the development of the questionnaire. All authors participated in the development process.

Funding The study was financed by the Norwegian Knowledge Centre for the Health Services (merged with the Norwegian Institute of Public Health in 2016).

Competing interests None declared.

Ethics approval The Norwegian Regional Committee for MedicalRegional committee for medical and health research ethics south east.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement The datasets generated and/or analysed during the current study are not publicly available due to protection of personal information but selected data are available from the corresponding author on reasonable request.

REFERENCES

1. Kelley E, Hurst J. Health care quality indicators project: conceptual framework. *OECD Health Working Papers, no 23*. Paris: OECD Publishing, 2006.
2. Murray CJ, Frenk J. A framework for assessing the performance of health systems. *Bull World Health Organ* 2000;78:717–31.
3. Anhang Price R, Elliott MN, Zaslavsky AM, *et al*. Examining the role of patient experience surveys in measuring health care quality. *Med Care Res Rev* 2014;71:522–54.
4. 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:e001570.
5. Donabedian A. The quality of care. How can it be assessed? *JAMA* 1988;260:1743–8.
6. Garratt AM, Bjertnaes OA, Barlinn J. Parent experiences of paediatric care (PEPC) questionnaire: reliability and validity following a national survey. *Acta Paediatr* 2007;96:246–52.
7. Garratt AM, Bjaertnes ØA, Krogstad U, *et al*. The OutPatient Experiences Questionnaire (OPEQ): data quality, reliability, and validity in patients attending 52 Norwegian hospitals. *Qual Saf Health Care* 2005;14:433–7.
8. Garratt AM, Danielsen K, Forland O, *et al*. The Patient Experiences Questionnaire for Out-of-Hours Care (PEQ-OHC): data quality, reliability, and validity. *Scand J Prim Health Care* 2010;28:95–101.
9. Garratt AM, Bjertnaes OA, Holmboe O, *et al*. Parent experiences questionnaire for outpatient child and adolescent mental health services (PEQ-CAMHS Outpatients): reliability and validity following a national survey. *Child Adolesc Psychiatry Ment Health* 2011;5:18. English.
10. Iversen HH, Holmboe O, Bjertnæs OA. The Cancer Patient Experiences Questionnaire (CPEQ): reliability and construct validity following a national survey to assess hospital cancer care from the patient perspective. *BMJ Open* 2012;2:e001437.
11. Skudal KE, Garratt AM, Eriksson B, *et al*. The Nordic Patient Experiences Questionnaire (NORPEQ): cross-national comparison of data quality, internal consistency and validity in four Nordic countries. *BMJ Open* 2012;2:e000864.
12. Danielsen K, Iversen H, Bjaertnes O. Pasienterfaringer med allmennleger–en kartlegging av relevante måleinstrumenter. *Notat fra Kunnskapssenteret november 2009*. Oslo: Nasjonalt kunnskapssenter for helsetjenesten, 2009.
13. Grol R, Wensing M. *Patients evaluate general/family practice: the EUROPEP instrument*. University Medical Centre: Nijmegen, 2000.
14. Steine S, Finset A, Laerum E, A new LE. A new, brief questionnaire (PEQ) developed in primary health care for measuring patients' experience of interaction, emotion and consultation outcome. *Fam Pract* 2001;18:410–8.
15. Bjertnaes OA, Lyngstad I, Malterud K, *et al*. The Norwegian EUROPEP questionnaire for patient evaluation of general practice: data quality, reliability and construct validity. *Fam Pract* 2011;28:342–9.
16. Vedsted P, Sokolowski I, Heje HN. Data quality and confirmatory factor analysis of the Danish EUROPEP questionnaire on patient evaluation of general practice. *Scand J Prim Health Care* 2008;26:174–80.
17. Holmboe O, Danielsen K, Iversen H. Utvikling av metode for måling av pasienters erfaringer med fastleger. *PasOpp-rapport*. Oslo: Norwegian Knowledge Centre for the Health Services, 2015.
18. Helsedirektoratet. Fastlegestatistikken. Directorate of Health. 2013 https://helsedirektoratet.no/statistikk-og-analyse/fastlegestatistikk#fastlegestatistikk-2013
19. Evans RG, Edwards A, Evans S, *et al*. Assessing the practising physician using patient surveys: a systematic review of instruments and feedback methods. *Fam Pract* 2007;24:117–27.
20. Campbell J, Smith P, Nissen S, *et al*. 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;10:57.
21. Lurås H. The association between patient shortage and patient satisfaction with general practitioners. *Scand J Prim Health Care* 2007;25:133–9.
22. Ramsay J, Campbell JL, Schroter S, *et al*. The General Practice Assessment Survey (GPAS): tests of data quality and measurement properties. *Fam Pract* 2000;17:372–9.
23. Hair J, Anderson R, Tatham R, *et al*. *Multivariate data analysis with readings*. 4 edn. Englewood Cliffs, New Jersey: Prentice-Hall International, 1995.
24. Mercer SW, Howie JG. CQI-2--a new measure of holistic interpersonal care in primary care consultations. *Br J Gen Pract* 2006;56:262–-68.
25. Nunnally J, Bernstein I. *Psychometric theory*. 3 ed. New York: McGraw Hill, 1994.
26. Gandek B, Ware JE, Aaronson NK, *et al*. Tests of data quality, scaling assumptions, and reliability of the SF-36 in eleven countries: results from the IQOLA Project. International Quality of Life Assessment. *J Clin Epidemiol* 1998;51:1149–58.
27. Sirdifield C, Godoy Caballero A, Windle K, *et al*. Comparing importance and performance from a patient perspective in English general practice: a cross-sectional survey. *Fam Pract* 2016;33:179–85.
28. Berendsen AJ, Groenier KH, de Jong GM, *et al*. Assessment of patient's experiences across the interface between primary and secondary care: Consumer Quality Index Continuum of care. *Patient Educ Couns* 2009;77:123–7.
29. Glasgow RE, Wagner EH, Schaefer J, *et al*. Development and validation of the Patient Assessment of Chronic Illness Care (PACIC). *Med Care* 2005;43:436–44.
30. King J, Gibbons E, Graham C, *et al*; *Developing measures of people's self-reported experiences of integrated care*. Picker Institute Europe and University of Oxford: Oxford, 2013.
31. Singer SJ, Friedberg MW, Kiang MV, *et al*. Development and preliminary validation of the Patient Perceptions of Integrated Care survey. *Med Care Res Rev* 2013;70:143–64.
32. Uijen AA, Schellevis FG, van den Bosch WJ, *et al*. Nijmegen Continuity Questionnaire: development and testing of a questionnaire that measures continuity of care. *J Clin Epidemiol* 2011;64:1391–9.
33. Haugum M, Danielsen K, Iversen HH, *et al*. The use of data from national and other large-scale user experience surveys in

local quality work: a systematic review. *Int J Qual Health Care* 2014;26:592–605.
34. Heje HN, Vedsted P, Olesen F. General practitioners' experience and benefits from patient evaluations. *BMC Fam Pract* 2011;12:116.
35. Iversen HH, Bjertnæs ØA, Groven G, *et al*. Usefulness of a national parent experience survey in quality improvement: views of paediatric department employees. *Qual Saf Health Care* 2010;19:e38.
36. Bjertnaes OA, Garratt A, Botten G. Nonresponse bias and cost-effectiveness in a Norwegian survey of family physicians. *Eval Health Prof* 2008;31:65–80.
37. Guldvog B, Hofoss D, Pettersen KI, *et al*. *PS-RESKVA -pasienttilfredshet i sykehus*. : Tidsskr Nor Laegeforen, 1998:118: 386–91.
38. Meterko M, Restuccia JD, Stoltzmann K, *et al*. Response rates, nonresponse bias, and data quality: results for a national survey of senior healthcare leaders. *Public Opin Q* 2015;79:130–44.
39. Hove G, Heje HN, Olesen F, *et al*. *DanPEP: Danske Patienter Evaluerer Almen Praksis; Projektafrapportering December 2002 - Marts 2006*. Danish: Århus: Forskningsenheden for Almen Praksis.
40. Roberts MJ, Campbell JL, Abel GA, *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:g6034.
41. Heje HN, Olesen F, Vedsted P. Patienternes vurdering af de praktiserende læger. *Overordnede resultater fra DanPEP-projektet*. : Ugeskr Laeger, 2010:172: 1105–12.
42. Nossen JP. Hva foregår på legekontorene? Konsultasjonsstatikk for 2006. *NAV-rapport 4-2007*. Oslo: Arbeids- og velferdsdirektoratet - Statistikk og utredning, 2007.