Involving patients in detecting quality gaps in a fragmented healthcare system: development of a questionnaire for Patients' Experiences Across Health Care Sectors (PEACS)

STEFAN NOEST[1], SABINE LUDT[1], ANJA KLINGENBERG[2], KATHARINA GLASSEN[1], FRIEDERIKE HEISS[1], DOMINIK OSE[1], JUSTINE ROCHON[3], KAYVAN BOZORGMEHR[1], MICHEL WENSING[4] AND JOACHIM SZECSENYI[1,2]

[1]Department of General Practice and Health Services Research, University Hospital Heidelberg, Heidelberg, Germany, [2]AQUA-Institute for Applied Quality Improvement and Research in Healthcare, Goettingen, Germany, [3]Institute of Medical Biometry and Informatics, University of Heidelberg, Heidelberg, Germany, and [4]Scientific Institute for Quality of Healthcare, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands

Address reprint requests to: Stefan Noest, Department of General Practice and Health Services Research, University Hospital Heidelberg, Vosstrasse 2, D-69115 Heidelberg, Germany. Tel: +49-6221-565660; Fax: +49-6221-561972; E-mail: stefan.noest@med.uni-heidelberg.de

Accepted for publication 16 March 2014

Abstract

Objective. The purpose of this study was to develop and validate a generic questionnaire to evaluate experiences and reported outcomes in patients who receive treatment across a range of healthcare sectors.

Design. Mixed-methods design including focus groups, pretests and field test.

Setting. The patient questionnaire was developed in the context of a nationwide program in Germany aimed at quality improvements across the healthcare sectors.

Participants. For the field test, 589 questionnaires were distributed to patients via 47 general practices.

Main Measurements. Descriptive item analyzes non-responder analysis and factor analysis (PCA). Retest coefficients (*r*) calculated by correlation of sum scores of PCA factors. Quality gaps were assessed by the proportion of responders choosing a response category defined as indicating shortcomings in quality of care.

Results. The conceptual phase showed good content validity. Four hundred and seventy-four patients who received a range of treatment across a range of sectors were included (response rate: 80.5%). Data analysis confirmed the construct, oriented to the patient care journey with a focus on transitions between healthcare sectors. Quality gaps were assessed for the topics 'Indication', including shared-decision-making (6 items, 24.5–62.9%) and 'Discharge and Transition' (10 items; 20.7–48.2%). Retest coefficients ranged from *r* = 0.671 until *r* = 0.855 and indicated good reliability. Low ratios of item-non-response (0.8–9.3%) confirmed a high acceptance by patients.

Conclusions. The number of patients with complex healthcare needs is increasing. Initiatives to expand quality assurance across organizational borders and healthcare sectors are therefore urgently needed. A validated questionnaire (called PEACS 1.0) is available to measure patients' experiences across healthcare sectors with a focus on quality improvement.

Keywords: quality measurement , quality management, patient satisfaction, measurement of quality, patient-centered care, quality improvement

Introduction

Measurement and reporting of patients' experiences have become an important element of health-service evaluation worldwide [1]. Several reports, in particular from the USA, have shown that fragmentation of modern healthcare systems has serious implications for patients and their quality of care [2], despite the stipulation that modern healthcare systems provide a sophisticated level of medical care, knowledge and technology. However, fragmentation between sectors also poses a challenge for patients and healthcare providers. This is most evident in the increasing complexity of managing patients living with chronic illness and cancer, where many actors are involved in the delivery of a complex chain of care across multiple sectors [3]. In turn, this leads to particular problems at the transition points between care sectors. Patients report problems with respect to coordination and cooperation between healthcare providers, with the consequence that problems in patient safety and quality of care arise [4, 5]. The current challenge posed is to expand quality assurance to include a cross-sectoral focus [3]. Patients' perspectives are critical to this expansion process, not only because patients are the only ones with first-hand experience of the different sectors from start to finish during their care journey, and therefore potentially offer a useful overview, but also because it is best practice to involve patients' perspectives in quality improvement initiatives [6, 7] as part of a move towards patient-centered care in modern health systems [2, 8].

In Germany, as in many other countries, quality assurance systems are typically restricted to either outpatient care or hospital care. However, a comprehensive program for quality improvement across healthcare sectors in Germany ('Sektoruebergreifende Qualitaetssicherung im Gesundheitswesen' or 'SQG'), established in 2009, is focused on the complete patient care journey across sectoral borders: from the phase of diagnosis to the phase of discharge and transition continued with the outpatient follow-up care [9]. As the SQG program is based on quality indicator development and measurement, feedback from patients via surveys provides an important method to evaluate performance of state-funded healthcare providers. To date, some topics for indicator development have included cataract surgery, cervical conization, colorectal cancer and percutaneous coronary interventions or coronary angiography [9].

International approaches and best practices in designing national quality improvement programs adopted multiple methods and steps for the development of patient questionnaires. One of the pioneers was the patient survey developed by the Consumer Assessment of Healthcare Providers and Systems Consortium (CAHPS) in the USA, which was used within a national quality program measuring patients' experiences of hospital care [10, 11]. The Netherlands adopted the CAHPS questionnaire as a component of their Consumer Quality Index (CQ index), a measurement to compare consumer experiences in health care for national health planning [12]. Numerous other countries use patient surveys for public reporting or health planning at a national level, e.g. the UK (NHS/Picker Institute Europe), Canada (Canadian Community Health Survey), Norway (Norwegian Knowledge Centre for the Health Services) or Denmark (Department of Quality Measurement for Aarhus) [1]. However, currently available instruments gathering and representing patients' views are either focused on organizational service development or they are limited to sector-specific contexts either to outpatient or in-patient care. To the authors' knowledge, no established tool is available to evaluate patients' experiences along their complete journey across healthcare sectors, which led to the development of this new instrument: PEACS 1.0 (Patients Experiences Across Care Sectors). We report the process of development and validation of this instrument, in a German context, to measure the quality of care across sectors from the patients' perspective.
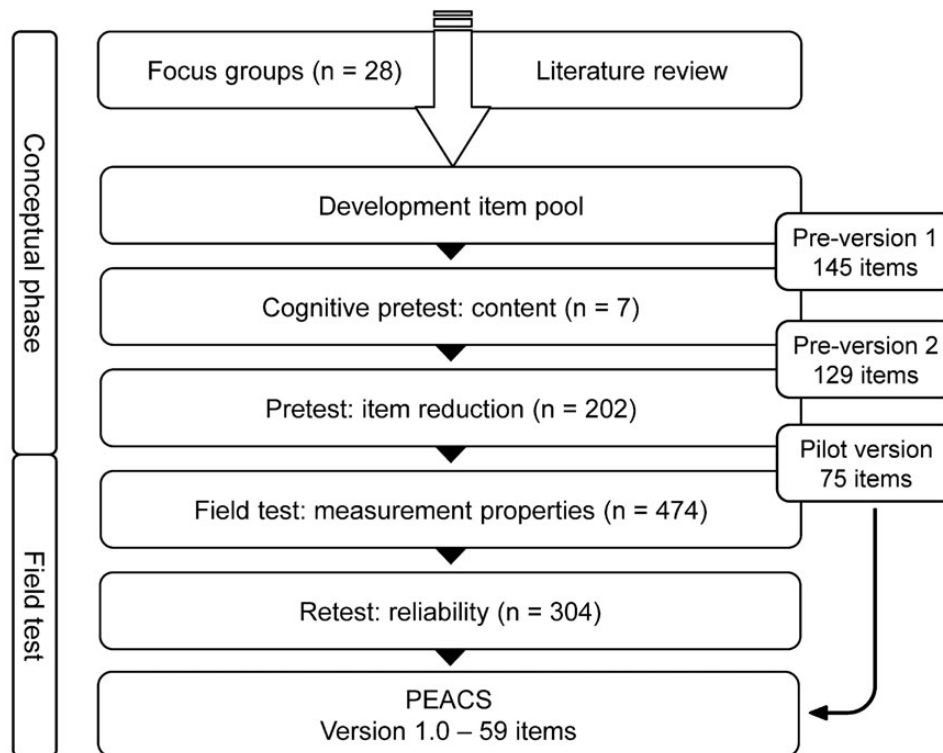
Method

Study design and setting

The instrument was developed between June 2011 and September 2012 in collaboration between the Department of General Practice and Health Services Research at University Hospital Heidelberg and the AQUA-Institute for Applied Quality Improvement and Research in Healthcare, Goettingen.

The mixed-method study design (Fig. 1) started with a conceptual phase including a qualitative focus group study to identify a broad range of patient perspectives [13], although in this paper, we primarily describe the quantitative field test. The goals of the field test were (a) to assess validity and reliability in accordance with the concept of measurement properties and proposed quality criterions for health-related patient-reported outcomes composed by the COSMIN initiative [14, 15] and (b) to determine whether the items were able to measure cross-sectoral quality. Before reporting the methods and results of the main field test, relevant background information related to the conceptual phase is described.

Conceptual phase

Step 1: focus groups. Qualitative data on patient perspectives were gathered via focus groups with a total of 28 patients. Following a literature review, a guide for the focus groups was created. The major subjects identified were communication, care-management, shared-decision-making, patient safety, patient support and frameworks. Focus group feedback was plotted along the cross-sectoral patient care journey to develop the questionnaire construct. Thus, the construct included a collection of major and minor subjects mapping a generic patient perspective onto the quality of cross-sectoral care. A core finding from the focus groups was that discharge and transition to home, or to follow-up care, were key areas where quality deficits occurred. Based on focus group data and the ensuing construct, an item pool of 145 questions was developed. The results of the focus-group study and the construct have been published separately [13].

Step 2: item design. Two central concepts for measuring patients' views are rating and reporting. There is a long tradition using rating scales for global measurement of patient

Figure 1 Mixed-method study design.

satisfaction [16–18]. Since the 1990s, evidence that reporting of experiences can be more helpful for quality improvement than global ratings of satisfaction has grown [19, 20]. In this study, reporting items, supplemented by rating items, and a 4-item scale for self-reporting outcome formed the conceptual basis of the questionnaire. The report questions were grouped into five composites and the rating items into three composites. The construct of the composites followed the process of cross-sectorial care including the following process phases: 'Diagnosis', 'Indication', 'Treatment at hospital/institution', 'Discharge and Transition' and 'Outpatient/Follow-up care'. The scales were not designed as homogenous subscales. Residual categories were added to the reporting items within the meaning of 'Was not important to me' or 'Does not apply to me'. For rating items, a five-point Likert scale was used ranging from 'fully correct' to 'not correct at all' also with one residual category.

Further information on Step 3 (Cognitive pretest) and Step 4 (Pretest and item reduction) of conceptual phase is summarized in Supplementary data, Appendix S1. In the following section, the main field test will be described.

Data collection

The pilot version of the questionnaire was administered to patients during a 2-week period in May 2012. Patients were recruited by 47 participating primary healthcare practices, which were selected by convenience sampling from an existing network of research practices [21]. Practice staff documented patient information from the recruited participants in a pseudonymized form.

In accordance with the SQG program and the core goals driving the development of a generic cross-sectorial questionnaire, the inclusion criteria for the recruiting patient participants were (a) involved in one of the target areas of the SQG program topics or (b) another surgery or treatment that had occurred in the previous 12 months, including both inpatient as well as outpatient settings (Table 1). In addition, patients had to be over 18 years. Patients were re-contacted 3 weeks after the first questionnaire was returned and asked to complete the questionnaire a second time for the retest results. Reminders or incentives for patients were not given.

Data analysis

Non-response analysis. A non-responder analysis was conducted to prevent a bias in interpreting data [22, 23]. A *t*-test was used for independent samples testing significance in differences of age and a cross-tabulation with chi-square tests was used for sex and treatment.

Descriptive item analyses. To describe sample and measurement characteristics, counts and percentages were used (Table 2). The acceptance of the instrument was based on the item non-response rate, calculated as the percentage of responders who did not provide valid responses for each item. The analysis of the residual category enables estimation of whether an item is important for the target group. Ceiling effects were also checked by evaluating the proportion of patients using the highest response category. Eight-five percent was applied, which is a common limit for ceiling effects [24], unless conceptual reasons

Table 1 Patient characteristics

	n	%
Age (n = 465[a])		
18–40 years	31	6.7
41–60 years	141	30.3
61–80 years	260	55.9
>80 years	33	7.1
Sex (n = 466[a])		
Female	242	51.9
Male	224	48.1
Treatment (n = 457[a])		
Topics of the SQG program	241	52.7
Conization (cervical portio)	3	0.7
Nosocomial infection	3	0.7
Colon cancer surgery	9	2.0
Breast cancer surgery	17	3.7
Knee replacement surgery	31	6.8
Knee arthroscopy	33	7.2
Hip replacement surgery	37	8.1
Cataract surgery	46	10.1
Percutaneous coronary intervention, cardiac catheter examination	62	13.6
Other surgery/treatment	216	47.3
Orthopedic surgery	88	19.3
Abdominal surgery	33	7.2
Thoracic/cardiovascular surgery	22	4.8
Other	73	16.0
Residence (n = 466[a])		
Rural	247	53.0
Small town	151	32.4
Urban	68	14.6
Partnership (n = 466[a])		
Single	104	22.3
Partnership	349	74.9
Other	13	2.8
Employment status (n = 461[a])		
Employed	154	33.4
Unemployed	10	2.2
Retired	277	60.1
Other	20	4.3
Health insurance (n = 456[a])		
Statutory	425	93.2
Private	31	6.8

[a]Differences to n = 474 are due to invalid or missing data.

were found. To measure quality, the particular response categories indicating low quality were defined for each item. We give a translated example for a reporting item with the indications of each response category:

Did the doctor talk to you about risks and possible complications of the treatment? (Item 14)

Yes, in detail (response defined as quality target)
Yes, partly (response defined as low quality)
No (response defined as low quality)

I did not want to talk about risks and complications (residual category)

In the case of rating items with a five-point Likert scale the answer 'fully correct' specifies the ceiling effect. The options 'partly', 'not correct' and 'not correct at all' were defined as low quality. The sixth category 'I don't know' was designed as a residual category. We calculate a quality gap measure per item as proportion of patient responses indicating low quality.

Dimensionality and construct validation. To examine the hidden domains of the questionnaire, a factor analysis (Principal Component Analyses or PCA) was conducted as a standard method [25]. We used the Kaiser–Meyer–Olkin measure (KMO) of sampling adequacy and the Bartlett test of sphericity to determine appropriateness of PCA. Two PCAs tests were run for reporting as well as rating items. The residual categories were recoded into missing values for PCA. Missing values were deleted pairwise. Oblique rotation (Promax) was used for reporting and Varimax for rating items.

Criterion validation. Sector transitions are the critical phases for quality problems in a typical care process across different sectors. Therefore, the questionnaire included items aiming to evaluate shared decision-making (Indication scale) and transition to follow-up care (Discharge and Transition scale). To assess criterion validity for these topics, the validated German version of the 9-item Shared Decision Making Questionnaire (SDM-Q-9) [26] and a self-translated version of the 3-item short form of Care-Transition-Measurement Questionnaire (CTM) [5] were included. The SDM-Q-9 measures the degree of SDM focusing the decision about different treatment options. On the other hand, the PEACS assesses SDM as information and involvement process independent from another available options. The CTM measured the overall quality of care transition assessing respect of patient's preferences (Item 1), understanding responsibility for self-managing health (Item 2) and understanding of medication purposes (Item 3). In contrast to the CTM-3, the items of PEACS place emphasis on communication of information in a more detailed way. Sum scores and correlated means tests comparing the external scale with our corresponding scales were also run.

Test–retest reliability. The test–retest method was applied to evaluate the validity and reproducibility of the survey. The Cohen's Kappa coefficient quantifies the intra-rater agreement per item. In consideration of the paradox characteristics of the coefficient [27], we decided for the weighted kappa k_w [28] with a weight of 0.75. We interpreted the coefficient in accordance with Altman and defined a limit of $k_W = 0.6$ [29]. If a value was <0.6 we reviewed the proportion of overall agreement and the agreement matrix. In the case of inconsistency, the item was removed. For measuring test–retest reliability at the level of construct, we calculated a summary index for each factor and correlated the means. Values >0.7 are usually regarded as confirming reliability and this was our cut-off point [30].

Data analyses were made using IBM SPSS Statistics version 20 except for the weighted kappa.

Table 2 Descriptive analyses and quality gap for items of PEACS 1.0 (field test: $n = 474$)

	No.	Factor	Item non-response n (%)	Residual category n (%)	Ceiling effect n (%)	Quality gap n (%)
Reporting items						
Diagnosis						
Physician in charge (nominal filter item)	1	–	30 (6.3)	–	–	
Information about results of initial examination	2	–	16 (3.4)	2 (0.4)	353 (77)	79 (17.4)
Indication						
Physician in charge (nominal filter item)	3	–	16 (3.4)	–	–	
Advice of a second opinion	4	I	16 (3.4)	280 (61.1)	117 (66)	61 (34.3)
Information about alternative treatment	5	I	13 (2.7)	268 (58.1)	116 (60)	77 (39.9)
Further information about disease	6	I	19 (4.0)	234 (51.4)	82 (37)	139 (62.9)
Possibility/consequence of non-perform treatment	7	I	20 (4.2)	209 (46.0)	162 (66)	83 (33.9)
Self-assessment participation	8	I	22 (4.6)	35 (7.7)	315 (76)	102 (24.5)
Treatment at hospital/institution						
Enough time for reading the informed consent form	11	II	4 (0.8)	33 (7.0)	383 (88)	54 (12.4)
Pre-operative consultation (filter item)	12	–	–	–	–	–
Explanation of the treatment process	13	II	8 (1.7)	6 (1.3)	371 (81)	89 (19.3)
Review risks of treatment	14	I	10 (2.1)	19 (4.1)	300 (67)	145 (32.6)
Possibility to ask questions	15	II	14 (2.9)	25 (5.4)	341 (78)	94 (21.6)
Satisfaction with answers	16	II	8 (1.7)	36 (7.7)	334 (78)	96 (22.3)
Understanding of patient briefing	17	II	11 (2.3)	8 (1.7)	344 (76)	111 (24.4)
Assistance through nurses	18	III	24 (5.1)	35 (7.8)	383 (92)	32 (7.7)
Information about state of health (nurses)	19	III	29 (6.1)	–	369 (83)	76 (17.1)
Nurses were responsive to fears and worries*	20	–	–	–	–	
Nurses were responsive to wishes and needs	21	III	15 (3.2)	196 (42.7)	213 (81)	44 (12.3)
Availability of physicians to address a question	22	IV	15 (3.2)	97 (21.1)	243 (67)	119 (32.9)
Information about state of health (physician)	23	II	12 (2.5)	–	394 (85)	68 (14.7)
Physician were responsive to fears/worries	24	–	9 (1.9)	269 (57.8)	144 (73)	51 (16.3)
Availability of physicians for relatives	25	IV	20 (4.2)	227 (50.0)	144 (63)	83 (36.6)
Waiting time before examination	26	–	9 (1.9)	20 (4.2)	–	259 (58.2)
Conflicting information by different persons	27	–	16 (3.4)	30 (6.6)	325 (76)	103 (24.1)
Prevent false-side mistake (surgery)	28	–	15 (3.2)	153 (33.3)	229 (75)	24 (9.5)
Medication error	29	–	12 (2.5)	47 (10.2)	405 (98)	10 (2.4)
Culture of dealing with error	30	–	21 (4.4)	366 (80.8)	39 (45)	48 (55.2)
Facility cleanliness	31	–	14 (2.9)	–	374 (81)	86 (18.7)
Adherence to hand hygiene (professionals)	32	–	10 (2.1)	105 (22.6)	314 (87)	45 (12.5)
Frequency of asking for pain	33	V	11 (2.3)	100 (21.6)	311 (86)	52 (14.3)
Immediate receive of pain therapy	34	V	33 (6.9)	165 (37.4)	248 (90)	28 (10.1)
Participation on pain treatment	35	–	33 (6.9)	235 (53.3)	147 (71)	59 (28.6)
Discharge and transition						
Information about (not) allowed behavior	36	VI	15 (3.2)	11 (2.4)	303 (68)	145 (32.4)
Information about danger signals	37	VI	11 (2.3)	76 (16.4)	242 (63)	145 (37.5)
Information about contact person and data	38	VI	16 (3.4)	21 (4.6)	336 (77)	101 (23.1)
Information on self-management	39	VI	18 (3.8)	82 (18.0)	240 (64)	134 (35.8)
Information on further treatment measures	40	VI	15 (3.2)	38 (8.3)	334 (79)	87 (20.7)
Information on course of disease	41	VI	23 (4.8)	24 (5.3)	302 (71)	125 (29.3)
Clarification of important issues	42	VI	15 (3.2)	89 (19.4)	274 (74)	96 (25.9)
Support by institution on transition to home	43	–	23 (4.8)	314 (69.6)	71 (52)	66 (48.2)
Involvement of relatives	44	–	13 (2.7)	299 (64.9)	121 (75)	41 (25.3)
Support in organization of follow-up rehabilitation	45	–	44 (9.3)	294 (68.4)	105 (77)	31 (22.8)
Written information for home care providers	46	–	41 (8.6)	406 (93.8)	17 (63)	10 (37.0)
Outpatient follow-up care						

(continued)

Table 2 Continued

	No.	Factor	Item non-response n (%)	Residual category n (%)	Ceiling effect n (%)	Quality gap n (%)
Continuity of pain treatment	51	–	25 (5.3)	202 (45.0)	192 (78)	55 (22.3)
Information about effects of medication	52	VI	16 (3.4)	107 (23.4)	262 (75)	89 (25.4)
Receive of written drug information	53	VI	19 (4.0)	123 (27.0)	250 (75)	82 (24.7)
Rating items						
Previously care						
Outpatient treatment	9	VII	20 (4.2)	4 (0.9)	342 (76)	46 (10)
Preliminary of medical intervention	10	VII	15 (3.2)	4 (0.9)	342 (75)	43 (9.5)
Hospital/institution						
Cooperation	47	IIX	14 (2.9)	33 (7.2)	299 (70)	44 (10.3)
Communication among professionals	48	IIX	14 (2.9)	21 (4.6)	308 (70)	38 (8.7)
Feeling of being in good hands	49	IIX	16 (3.4)	0 (0.0)	358 (78)	38 (8.3)
Recommendation to friends/relatives	50	IIX	11 (2.3)	2 (0.4)	332 (72)	53 (11.5)
Transition to follow-up care or home						
Smoothless success of transition	54	IIX	29 (6.1)	35 (7.9)	314 (77)	37 (9.0)
Provider taking care for follow-up care	55	IIX	36 (7.6)	88 (20.1)	249 (71)	52 (14.9)
Outcome						
Improvement of disorders	56	IX	15 (3.2)	65 (14.2)	241 (61)	57 (14.5)
Complications due to treatment	57	IX	18 (3.8)	–	349 (77)	107 (23.5)
Satisfaction with outcome	58	IX	14 (2.9)	8 (1.7)	290 (64)	85 (18.8)
Decision for intervention was all right	59	IX	18 (3.8)	12 (2.6)	355 (80)	37 (8.3)

[a]Item added after field test.

Ethics

Ethical approval for this study was received from the Ethics Committee of the Medical Faculty at the University of Heidelberg (Reference: S-586/2011).

Results

We received 492 responses and excluded 16 forms because of missing or invalid data. Two patients refused participation (final response rate: 474/589, 80.5%). For the retest phase, 32 participants were lost to the study due to absent or false address data or a decision to withdraw. Four hundred and forty-two participants were included for the retest and from this remaining pool, 342 questionnaires were returned (77.3%). We reviewed the questionnaire for sufficient data and plausibility (concordance of treatment data between t_0 and t_1) and excluded 38 further questionnaires (final response rate retest: 304/442, 68.8%) (Fig. 2).

Study population

Table 1 provides socio-demographic information and other characteristics of the patient sample. Approximately half of participants were female (51.9%). The mean age was 63.2 years (SD 14.5). Health care for most participants was funded by the statutory German national health insurance scheme (93.2%).

Non-responder analysis

The mean age of non-responders was 60.6 (SD 17.5) years and 49.1% ($n = 53$) were female. The distribution of treatment groups was also similar in both samples. Responders did not significantly differ from non-responding patients in these relevant characteristics: age ($P = 0.11$), sex ($P = 0.59$) and treatment groups ($P = 0.42$).

Measurement properties

Table 2 provides descriptive information on the items of PEACS 1.0.

Descriptive item analysis. The proportion of missing values at item level ranged from 0.8 to 9.3%. Item 45 was developed as a filter question. Overall, 47 of 57 items had a non-response rate <5%. Ceiling effects ranged from 37 to 98%. Item 29 measured a rare event, which may have serious consequences for patients. Some items exhibit a high ratio of participants choosing the residual category, e.g. item 30 (80.8%, $n = 366$) or item 46 (93.8%, $n = 406$). The proportion of answers indicating quality gaps by reporting items ranged from 2.4 to 58.2%. A concentration of quality gaps at the phases of the continuum of care *Indication* and *Discharge and Transition* was observed. Quality gaps indicated by rating items were relatively moderate (8.3–23.5%) in comparison with reporting items, with an enhanced concentration on self-reported *Outcome*.
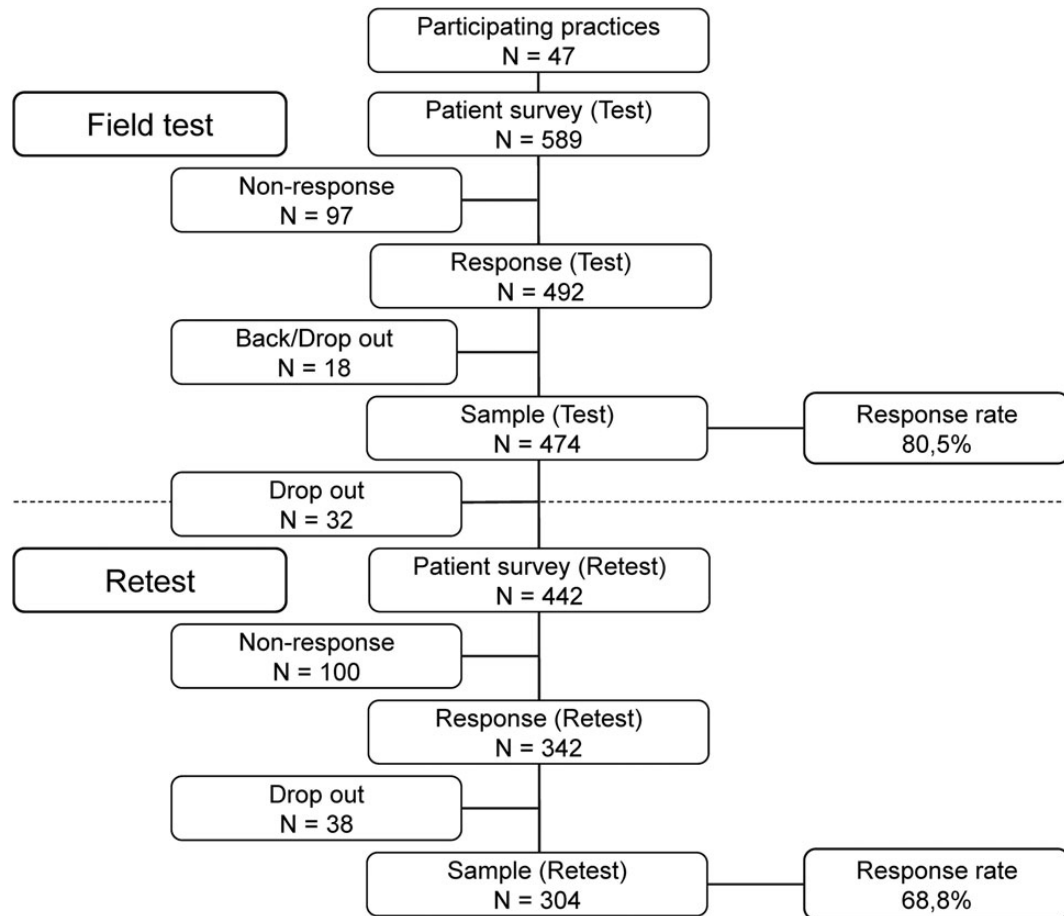
Figure 2 Sample size and response rate for field test and retest.

Dimensionality and construct validation. Some reporting items were excluded for PCA step-by-step analysis because the amount of residual category (missing values for factor analysis) was high or Measure of Sample Adequacy (MSA) was low (e.g. Item 27; MSA = 0.08). Items were also excluded if they loaded on more than one factor (e.g. Item 51). The final PCA included 28 reporting items with different sample sizes each (mean 370, SD 87, min 128, max 462 participants). The appropriateness of the items was very good (KMO: 0.891, Bartlett significance: 0.000). Based on the Kaiser criterion (Eigenvalue>1) items are categorized in six different factors explaining 64.0% of the total variance between the 28 variables. Additionally, we conducted a second PCA and included all rating items with very good appropriateness (KMO 0.83, Bartlett significance 0.000) resulting. Items were categorized into three different factors explaining 69.8% of the total variance between the 12 variables. Factor loadings of PCAs are shown in Supplementary data, Appendix S2.

Criterion validation. We found a significant and high correlation of the factor scale 'Shared decision-making at indication' and the SDM-Q-9 scale ($r = 0.814$, $P < 0.001$). Between the factor scale 'Information at discharge and follow-up' and the CTM-3 we assessed a significant and moderate correlation ($r = 0.511$, $P < 0.001$).

Retest reliability. All values for assessing reliability of PEACS 1.0 items are shown in Supplementary data, Appendix S3. Thirty-two items offered a good weighted kappa ($k_w > 0.6$). Eight items had a lower but acceptable weighted kappa (0.52–0.59) with good proportions of overall agreement (p_o) and concurrently no abnormal agreement matrix. Three items (23, 27 and 30) were conspicuous. Because of the importance of the item content, we decided to keep them into the instrument for a broader field test.

Test–retest correlations based on construct level are shown in Table 3. The retest coefficients indicated good reliability ($r > 0.7$), except for factor IIX with a moderate value ($r = 0.671$).

Overall, 16 items of the pilot version of the questionnaire were excluded to develop PEACS 1.0 (Supplementary data, Appendix S4).

Discussion

There is broad political and academic consensus supporting the measurement and improvement of quality across sectors, but the existing borders between the different sectors make it challenging [9]. In this study, we developed and validated a generic German questionnaire (PEACS 1.0) to evaluate

Table 3 Retest reliability construct level (based on PCA factors)

No.	Dimension	Number of items (n)	Retest coefficient (r)
Factors: Reporting items			
Factor I	Shared decision-making at indication	(6)	0.855
Factor II	Patient education and information	(6)	0.778
Factor III	Nursing staff	(3)	0.743
Factor IV	Availability physicians	(2)	0.717
Factor V	Pain therapy	(2)	0.741
Factor VI	Information at discharge and follow-up	(9)	0.763
Factors: Rating items			
Factor VII	Preliminary care	(2)	0.700
Factor IIX	Institutional treatment and transition	(6)	0.671
Factor IX	Outcome	(4)	0.803

patients' experiences and detect potential quality gaps along the complete journey of care. The stepwise development process supported good content validity. The high response rate and the very low item-non-response indicate a very high acceptance by patients. Reliability was considered to be good using the test–retest procedure. The moderate criterion correlation with the CTM-3 could be caused by the differences in the item contents. CTM-3 focused on the clinical level, while the factor scale 'information at discharge and follow-up' addresses clinical as well as outpatient levels. This example indicates one of the challenges in developing an appropriate instrument to measure care across sectors: methodological standards and good practices are still in the early development stage.

The need for integrating patients' perspectives in cross-sectoral quality improvement

Including patients' perspectives is fundamental to quality improvement in health care [2, 7]. Particularly in fragmented healthcare systems, patients are the only ones with first-hand experience of the different sectors from start to finish during their care journey. Therefore, an important part of our instrument development strategy included the involvement technique to identify patients' perspective and their preferences, using focus groups [31]. From the patients' perspective, processes of communication, coordination and transition were defined as relevant quality dimensions, sometimes without reference to the outcome of these processes. These results are accompanied by further evidence that patients wish to be informed and involved in care processes, whether on an individual basis to a greater or lesser extent [32]. We evaluated patients' feedback as an important source of information, even in the context of complex issues like patient safety and infection control [13], and even though the patient role in quality assurance remains controversial [33]. Several studies have shown, however, that patients are able to identify important care-related issues [34].

Evaluate quality gaps

Patient-reported experiences measured by reporting items demonstrated a wide range of quality gaps. This comes with the limitation that the ratios of quality gaps hinged on the number of patients who chose a response category, defined as problem, in relation to all patients who chose a response category that was different from the residual category. The advantage of offering a residual category is to exclude patients who are not concerned with a particular item as well as to evaluate the importance of item content. The disadvantage is the reduction of valid responses for data analysis [35]. Responses of item 30 ('Culture of dealing with errors'), for example, indicated that a high proportion of participants chose the residual category 'An error did not occur'. The information offered by the residual categories is worth being analyzed on its own: in the case of item 30, the first result is whether an error occurred or not (based on the residual category), and the second result is how patients experienced the situation was dealt with if an error occurred. In the case of Item 46 ('Written Information for home care providers' with the residual category 'Home care provider were not needed after transition'), for example, we measure a rare event that has a high potential for detecting quality gaps affecting a small-sized target group. We presume that this will be important in the case of a broader sample. Due to these analytical options and because of its importance for cross-sectoral quality of care, we decided to keep items in the questionnaire despite high residual categories. The decision for keeping or removing of items cannot only be determined in light of measurement properties. Overall, the ratios of quality gaps show that PEACS 1.0 is able to evaluate potential for quality improvement, even with the limitation of small samples for some items.

Application of the PEACS questionnaire

The developed PEACS 1.0 contains 59 items, based on the typical care process and focusing on transition processes. Domains in this version are 'Preliminary care', 'Shared decision-making at indication', 'Patient education and information', 'Nursing staff', 'Accessible physicians', 'Pain therapy', 'Institutional treatment and transition', 'Information at discharge and follow-up' and self-reported 'Outcome'. Every item in these domains indicated a quality target with possible problems. Because the scales are not designed homogenously,

the decision to choose or exclude an item had to be decided with relevance to quality assurance needs based on item content. To communicate the results of the assessment, e.g. to providers or to the public, it seemed to be meaningful to aggregate the item results as problem scores to topics. The PCA and statistical analysis identified relevant items to be aggregated in respective dimensions. This approach supported the initial construct of the questionnaire and confirmed the general construct of the patient care journey. However, further research for construct validation is necessary before reporting aggregated measures in a national program.

Some items had to be excluded for PCA because of statistical requirements, although they are important for quality assurance due to content, based on the conceptual phase. The single items 29, 30, 31, 32 ('Medication error', 'Culture of dealing with error', 'Facility cleanliness', 'Adherence to hand hygiene') relate to patient safety and the single items 43, 44, 45, 46 ('Support by institution on transition to home', 'Involvement of relatives', 'Support in organization of follow-up rehabilitation', 'Written information for home care providers') relate to important transition elements. These items refer to rare or adverse events (reason for high rate of residual category = missing value). They are very important for patients as evaluated in the conceptual phase, which is why we decided to keep them in the questionnaire even if they are not suitable for PCA.

The result of the PCA, shown in Supplementary data, Appendix S2, includes 40 items. We recommended these 40 items as a minimal generic set of the bank of items of PEACS 1.0 to assess patients' experiences of care across sectors. This 40-item version of PEACS questionnaire provides a statistically well-grounded assessment of quality gaps with a focus on the transition between sectors.

In addition to the PCA-based 40 item version, PEACS 1.0 includes 19 further items, e.g. filter items and items they are not suitable for PCA because of the high rate of residual category, but with important content from patient perspectives and a high potential of discriminative power. These items representing important patient experiences analyzed in focus groups have a high potential to discriminate healthcare provider based on quality measure, but with the limitation that the questionnaire PEACS 1.0 has to be tested on a bigger generic sample for final optimizing before applying it in the public SQG program. With this in mind, our study included patients with the index disease defined by the legislation topics of the SQG program. However, as indicated by Table 1, the study included a high proportion (47%) of patients with diseases beyond those SQG topics. We therefore consider the questionnaire to be applicable for a broad cross-section of patients.

Limitations

We constructed the PEACS 1.0 with an emphasis on quality of care processes. It must be stressed that the questionnaire was developed as a generic modular tool that is intended to be supplemented by specific topic-related items in patient surveys for the topics of the SQG program. The appropriateness of the instrument has to be evaluated in the light of this context.

Conclusions

Our study showed that it is possible to develop and to validate a survey instrument to detect quality gaps in fragmented health care by evaluating patients' experiences across different healthcare sectors. For benchmarking purposes and monitoring performance of healthcare providers over a period of time, it is necessary to validate newly developed questionnaires and test the discriminative power at the level of provider cluster. Indication-specific scales and disease-specific outcome measures will be added in a next step.

Supplementary material

Supplementary material is available at *INTQHC Journal* online.

Acknowledgements

Conflict of interests

J.S. is a director and shareholder of the AQUA-Institute which is responsible for the development of the SQG program in contract with the Federal Joint Committee. A.K. is an employee of the AQUA-Institute. The department of General Practice and Health Services Research (S.N., S.L., D.O., K.G., K.B.), the Institute of Medical Biometry and Informatics (J.R.) and the Radboud University Nijmegen (M.W.) are scientific partners in the SQG program.

Funding

References

1. Garratt A, Solheim E, Danielsen K. *National and cross-national surveys of patient experiences: a structured review* (Report No. 7). Oslo: Norwegian Knowledge Centre for the Health Services, 2008.

2. Institute of Medicine. *Crossing the Qualiy Chasm: A new Health System for the 21st Century.* Washington, DC: National Academy Press, 2001.

3. Sachverstaendigenrat zur Begutachtung der Entwicklung im Gesundheitswesen. *Special Report 2012. Competition at the interfaces between outpatient and inpatient care* (in German). Bern: Hans Huber, 2012.

4. Coleman EA, Berenson RA. Lost in transition: challenges and opportunities for improving the quality of transitional care. *Ann Intern Med* 2004;**141**:533–6.

5. Coleman EA, Mahoney E, Parry C. Assessing the quality of preparation for posthospital care from the patient's perspective—the care transitions measure. *Med Care* 2005;**43**:246–55.

6. Black N, Jenkinson C. Measuring patients' experiences and outcomes. *BMJ* 2009;**339**:b2495.

7. Wensing M, Elwyn G. Research on patients' views in the evaluation and improvement of quality of care. *Qual Saf Health Care* 2002;**11**:153–7.

8. Arah OA, Westert GP, Hurst J *et al*. A conceptual framework for the OECD Health Care Quality Indicators Project. *Int J Qual Health Care* 2006;**18**:5–13.

9. Szecsenyi J, Broge B, Eckhardt J *et al*. Tearing down walls: opening the border between hospital and ambulatory care for quality improvement in Germany. *Int J Qual Health Care* 2012;**24**:101–4.

10. Hays RD, Shaul JA, Williams VS *et al*. Psychometric properties of the CAHPS 1.0 survey measures. Consumer Assessment of Health Plans Study. *Med Care* 1999;**37**:MS22–31.

11. Lee Hargraves J, Hays RD, Cleary PD. Psychometric properties of the Consumer Assessment of Health Plans Study (CAHPS®) 2.0 Adult Core Survey. *Health Serv Res* 2003;**38**:1509–28.

12. Delnoij DMJ, Asbroek Gt, Arah OA *et al*. Made in the USA: the import of American Consumer Assessment of Health Plan Surveys (CAHPS®) into the Dutch social insurance system. *Eur J Public Health* 2006;**16**:652–9.

13. Ludt S, Heiss F, Glassen K *et al*. Patients' perspectives beyond sectoral borders between inpatient and outpatient care—patients' experiences and preferences along cross-sectoral episodes of care (in German). *Gesundheitswesen* 2013 (EFirst: 18 July 2013).

14. Mokkink LB, Terwee CB, Patrick DL *et al*. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;**63**:737–45.

15. Terwee CB, Bot SDM, de Boer MR *et al*. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;**60**:34–42.

16. Ware JE, Jr, Davies-Avery A, Stewart AL. *The Measurement and Meaning of Patient Satisfaction. A Review of the Literature*. Santa Monica, CA: The Rand Corporation, 1977.

17. Sitzia J. How valid and reliable are patient satisfaction data? An analysis of 195 studies. *Int J Qual Health Care* 1999;**11**:319–28.

18. Fenton JJ, Jerant AF, Bertakis KD *et al*. The cost of satisfaction a national study of patient satisfaction, health care utilization, expenditures, and mortality. *Arch Intern Med* 2012;**172**:405–11.

19. Cleary PD, Edgmanlevitan S, Roberts M *et al*. Patients evaluate their hospital-care—a national survey. *Health Aff* 1991;**10**:254–67.

20. Delbanco TL. Quality of care through the patient's eyes. *BMJ* 1996;**313**:832–3.

21. Peters-Klimm F, Freund T, Bentner M *et al*. In the practice and for the practice (in German). *German J Fam Med* 2013;**89**:183–8.

22. Edwards Philip J, Roberts I, Clarke Mike J *et al*. Methods to increase response to postal and electronic questionnaires. *Cochrane Database Syst Rev* 2009. Issue 4. Art. No.: MR000008.

23. Kelley K, Clark B, Brown V *et al*. Good practice in the conduct and reporting of survey research. *Int J Qual Health Care* 2003;**15**:261–6.

24. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual of Life Res* 1995;**4**:293–307.

25. Keller S, O'Malley AJ, Hays RD *et al*. Methods used to streamline the CAHPS Hospital Survey. *Health Serv Res* 2005;**40**:2057–77.

26. Kriston L, Scholl I, Holzel L *et al*. The 9-item Shared Decision Making Questionnaire (SDM-Q-9). Development and psychometric properties in a primary care sample. *Patient Educ Couns* 2010;**80**:94–9.

27. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. *J Clin Epidemiol* 1990;**43**:543–9.

28. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull* 1968;**70**:213–20.

29. Altman DG. *Practical Statistics for Medical Research*. London: Chapman and Hall, 1995.

30. Nunnally JC. *Psychometric Theory*. New York: McGraw-Hill, 1978.

31. Sixma HJ, Kerssens JJ, Campen Cv *et al*. Quality of care from the patients' perspective: from theoretical concept to a new measuring instrument. *Health Expect* 1998;**1**:82–95.

32. Guadagnoli E, Ward P. Patient participation in decision-making. *Soc Sci Med* 1998;**47**:329–39.

33. Armstrong N, Herbert G, Aveling EL *et al*. Optimizing patient involvement in quality improvement. *Health Expect* 2013;**16**:e36–47.

34. Zhu J, Stuver SO, Epstein AM *et al*. Can we rely on patients' reports of adverse events? *Med Care* 2011;**49**:948–55.

35. Porst R. Questionnaire: A Textbook *(in German)*. Wiesbaden: VS Verlag, 2011.