PEC Innovation 2 (2023) 100174

Contents lists available at ScienceDirect

PEC Innovation

journal homepage: www.elsevier.com/locate/pecinn

Russian-language translation and cultural adaptation of the Norwegian 'Patient Experience Questionnaire'

Mikhail Ionov [a,*,1], Elena Dubinina [b,c,1], Ilya Tregubenko [d,e,1], Nadezhda Zvartau [a,1], Alexandra Konradi [a,1]

[a] Research Laboratory on Pathogenesis and Therapy of Arterial Hypertension, Almazov National Medical Research Center of Ministry of Health of the Russian Federation, Saint-Petersburg, Russian Federation
[b] Department of Clinical Psychology and Psychological Assistance, Russian State Pedagogical University of Russia, Saint-Petersburg, Russian Federation
[c] Laboratory of Clinical Psychology and Psychodiagnostics, Bekhterev National Research Medical Center for Psychiatry and Neurology of Ministry of Health of the Russian Federation, Saint-Petersburg, Russian Federation
[d] Department of Psychology of Professional Activity and Information Technologies in Education, Russian State Pedagogical University of Russia, Saint-Petersburg, Russian Federation
[e] Department of General and Clinical Psychology, Pavlov First St.Petersburg State Medical University, Saint-Petersburg, Russian Federation

ARTICLE INFO

Keywords:
Surveys and questionnaires
Outpatient
Patient satisfaction
Validation studies
Translating
Cross-cultural comparison
Patient-centered care

ABSTRACT

The availability of patient-reported experience measures (PREM) is an unmet need in Russian healthcare.
Objective: To translate, adapt culturally, and validate PREM for outpatients.
Methods: A core set of questions from the Patient Experience Questionnaire (PEQ, in Norwegian, available in English) was translated to Russian (forward-backward translation). Acceptability, construct validity, and reliability were assessed. Patients aged ≥18 y.o. were invited to complete the questionnaire via QR-code within 24 h after a medical encounter.
Results: A questionnaire with adequate conceptual and linguistic equivalence was obtained. For four questions, a rating scale was replaced by Likert-type. A total of 308 responses were received (median age 55 y.o., 52% females). The correlation matrix was factorable. Four factors were extracted using varimax rotation: 1) outcome of this specific visit; 2) communication experiences; 3) communication competency; 4) emotions after this visit. These explained 65.4% of the total variance. Three items were excluded. The model was confirmed to be adequate. The Cronbach alpha was >0.9. Item-total correlation confirmed discriminative validity.
Conclusion: These preliminary results show that the Russian version of PEQ, adapted to national features, shows good psychometric properties. External validation is needed for the broad implementation of this PREM.
Innovation: This research is first attempt to use PREM in the Russian Federation. The use of quick response codes is feasible and eases survey conduction. The more PREMs are used the higher the quality of healthcare.

1. Introduction

With the advent of the Millennium, industrialized countries have faced the problem of a shortage of funds superimposed on social imbalances. A novel, value-based approach was proposed as one of the responses to these challenges [1]. Within the new paradigm, the patient 'values' have become a pivotal component in linking the clinical and financial components of the system. Aside from patient-reported outcomes (PROs), reflecting patient-perceived usefulness of interventions, the quality of care have become another central element [2]. Healthcare quality is commonly defined as 'a set of characteristics of medical care, reflecting its ability to meet the needs of patients, considering health standards that correspond to the current level of medical science' [3].

To objectively assess patients' needs, expectations, and experience with healthcare, special instruments are used called patient-reported experience measures or PREMs [4]. Patient experience assesses whether something that should happen in a healthcare setting (i.e., the proper communication with a doctor) actually happened or how often it happened [5]. These tools, like PROMs, may evaluate a variety of domains, including but not limited to engagement in the decision-making process, accessibility of services, and coordination of care. PREMs are also increasingly included in public reporting, benchmarking, and pay-for-performance programs [6,7]. For example, the higher the hospital scores in PREMs, the better treatment results and safety [8]. There is also a straight association between PROMs and PREMs. Black et al. have shown a 3% improvement in PROs with a concomitant 10% PRE increase [9].

Abbreviations: PRE(M), patient-reported experience (measures); PRO(M), patient-reported outcome (measures); KMO, Kaiser-Meyer-Olkin [test]; MCAR, missing completely at random; QR, quick response [code]; EFA, exploratory factor analysis; pCFA, partial confirmatory factor analysis.

* Corresponding author at: Almazov National Medical Research Center of Ministry of Health of the Russian Federation, 2 Akkuratova street, Saint-Petersburg 197341, Russian Federation.
E-mail address: ionov_mv@almazovcentre.ru (M. Ionov).

[1] All authors contributed to critical revision of the manuscript and approved the final version of the manuscript.

http://dx.doi.org/10.1016/j.pecinn.2023.100174
Received 4 December 2022; Received in revised form 25 May 2023; Accepted 2 June 2023

Several PREMs have been developed and successfully validated across Europe and the USA for in- and outpatient care [10]. However, in the Russian Federation, no PREMs exist. Therefore, healthcare benchmarking is often chaotic and only occasionally incorporates patient-perceived values. As per Federal Law #323-FZ 'On Basics of Health Protection of the Citizens in the Russian Federation' an independent assessment of the quality of medical services provided includes such parameters as: 'openness and accessibility of information, the comfort of conditions, waiting time, attitude of employees of the organization and their competence' [11]. This definition might be interpreted rather broadly, and quality of care analysis is carried out periodically by the judge since the information is provided mainly by the medical authorities. Several surveys were developed by the Ministry of Health of the Russian Federation [12], but they measure *satisfaction* with care and facility rather than patient-perceived *experience*. PREMs differ from satisfaction surveys by reporting objective patient experiences and removing subjectivity [4]. Due to, e.g., gratitude bias, patients frequently overrate their level of satisfaction. As a result, satisfaction questionnaires have certain limitations regarding reliability and utility [13].

Stressing the lack of valid and reliable PREMs in Russia, we aimed to validate one questionnaire developed abroad for ambulatory patients. Before being used in a new cultural or linguistic context, any questionnaire should be translated and appropriately adapted to be interpretable and to allow comparability between different contexts [14].

We picked the Patient Experience Questionnaire which provides experience and emotions after outpatient appointment. It was originally developed in Norway, and available in English language [15].

2. Methods

2.1. Study design

This was a cross-cultural adaptation of the Patient Experience Questionnaire (PEQ) and cross-sectional validation study.

This original questionnaire comprises four domains: 'Outcome of this specific visit' (4 items), 'Communication experiences' (4 items), 'Communication barriers' (4 items), 'Experience with the auxiliary staff' (2 items) and 'Emotions immediately after the visit' (4 items). The first 4 domains (14 items) are measured on a 5-point Likert scale ranging from 1 (worst-case scenario) to 5 (best-case scenario). The last emotional domain represents 4 emotions regarding one open-type statement ('After this visit, I felt:') with 7-point rating scale (again, from 1 to 7, meaning worst- and best-case scenarios, respectively).

It is the first questionnaire developed in a Scandinavian primary care setting and one of the first consultation-specific questionnaires not limited to satisfaction items.

The original article describes this questionnaire's conceptual basis and design [15]. Briefly, the development of the instrument involved consultation with experts, a systematic literature review, organization of patient focus groups, and in-depth interviews to determine issues of salience to them in health care encounters.

This study comprised several consequent stages: questionnaire re-development (forward-backward translations, cognitive debriefing, editing), mass survey, and internal validation per se (including exploratory and partial confirmatory factor analyses).

2.2. Questionnaire re-development

The authors-developers of PEQ were asked permission to use the tool. The English version of the questionnaire was obtained from the published article [15] and underwent a cross-cultural adaptation process following the guidelines established by the International Society for Pharmacoeconomics and Outcome Research [16] to obtain Russian version. These included investigation of conceptual and item equivalence, two English-Russian original PEQ forward translations and two Russian-English back translations, synthesis of the documents, and revision by an expert committee (two psychologists, two physicians). An expert panel confirmed linguistic equivalence at word and phrase level, as well as grammatical and pragmatic equivalence [17]. The conceptual equivalence was deemed satisfactory as our research team identified the exact five domains as the researchers in the original paper.

The cognitive debriefing was performed to test the version obtained. In July 2021, 13 ambulatory patients (6 male and 7 female patients) and 3 non-medical specialists working in the Centre were asked to complete the questionnaire. The mean completion time in non-medical staff was 3 min (1 to 5 min). Afterward, they were asked about how they had understood it. The participants needed help understanding the last part of the questionnaire as they most often contoured the extreme terms placed on the edges of the rating line. Though they were properly trained before going through the questionnaire, when they were inquired, eight patients answered that they could not match their emotions with numbers on the rating scale accordingly. Therefore, following a discussion with the pilot group of respondents, experts decided to replace the original rating scale with Likert-type answers for these four last questions retaining their initial meaning. Following this editing, patients from the pilot group became more confident in their answers when the study team interviewed them. The modified questionnaire was ready for a cross-sectional validation study.

2.3. Study participants

The study was conducted in the Diagnostic Center, an ambulatory division of Almazov National Medical Research Centre in Saint-Petersburg, Russia. Patients here receive high-quality counseling, various examinations, and long-term follow-up. It is a multispecialty clinic of >50 doctors serving patients within 33 medical specialties. A mean of 14,149 ambulatory appointments per month is performed (a total of ~170,000 patients per year from more than four regions, including Saint-Petersburg).

Fifteen doctors were randomly (sealed envelopes with names) selected. At the bottom of the printed medical certificates (a report given to a patient containing one's diagnosis and recommendations signed by an attending physician), a unique Quick Response (QR) code was generated automatically. The Internet link was embedded in the QR code, allowing online access to the questionnaire, and storing data anonymously in the internal server of the Centre. These QR codes could be activated only from medical certificates that randomly selected doctors gave. Codes from other doctors were inactive. QR codes were not associated with the patient's data. It allowed only one entry per person 24 h after an appointment. Patients were informed about the study objectives. Doctors obtained verbal (at the appointment) and electronic (following the link in the QR-code) informed consents from patients. Ethical approval for this study was sought and obtained from the Local Ethics Committee (approval No 77 dated May 15, 2017).

The patients were self-included as they proceeded to questionnaire polling via QR-code, so nonspecific inclusion/exclusion criteria were present. The questionnaire was polled only among adult patients (aged ≥18 years old); therefore, no pediatric patients were included in this study.

2.4. Statistical analysis

Descriptive statistics included frequencies and percentages of the categorical variables and means and standard deviation, or median with first and third quartiles, of the continuous variables. To identify possible differences between patient groups from the original paper and our research, we conducted the bivariate analysis using the chi-squared test for categorical variables and Manna-Whitney or Kruskal-Wallis tests for continuous variables. Questionnaire acceptability was judged according to the total percentage of respondents and the total completion time.

We applied an expectation maximization algorithm within the Little's Missing Completely at Random (MCAR) test to report the missing data's quantity and nature. The imputation technique was applied depending on the percentage of missing data and the results of the MCAR test [18]. The basic rule of thumb is that if <5% of the observations are missing, the missing data can be deleted without significant ramifications [19]. Based on the original study's results and since ours was a cross-sectional survey within a

smaller population than in the original research, we expected that far >5% of the answers would be missed. We employed a multiple imputation with regression [20].

Exploratory factor analysis was carried out to test construct validity. A correlation matrix was calculated. Bartlett's test of sphericity [21] first measured the strength of inter-correlation between items and tested whether the population correlation matrix was an identity matrix.

Kaiser-Meier-Olkin statistic was required to be above a minimum of 0.50 [22]. Factor extraction was performed using the minimum residual (MINRES) procedure with promax (oblique) and varimax rotations. The best and the simplest structure was selected according to the results of two rotations. The Parallel Analysis [23] and Cattel's visual scree test [24] were used to determine the appropriate number of factors to retain. Factors with an eigenvalue >1 were retained. Criteria for determining factor adequacy were established a priori. Given the number of participants in this study, pattern coefficients (factor loading) ≥0.30 might have been considered salient (as per Norman & Streiner [25]). However, we decided the extracted factors should be interpreted considering loading of 0.5 or above to indicate of the underlying dimension (based on the original research and [26]). We performed partial confirmatory factor analysis (pCFA) to support the results of the EFA at this stage of the study. The fit of the model was evaluated by the indexes as follows: standardized root mean square residual (SRMR) is <0.10; root mean square error of approximation (RMSEA) is <0.09; Tucker Lewis index (TLI) and comparative fit index (CFI) both are >0.90.

Questionnaire internal consistency was tested using Cronbach's alpha and McDonald's omega coefficients for each factor, as well as for the whole scale and after the suppression of each item. Factors with a minimum internal consistency reliability ≥0.70 and theoretically meaningful were considered adequate [27]. Finally, item-domain correlations were evaluated expecting moderate correlations since they assess a construct but are not considered paraphrases.

Statistical analyses were performed using SPSS software Version 23.0 (IBM, Armonk, NY, USA) and jamovi. (the jamovi project, version 1.6 for MacOS, retrieved from https://www.jamovi.org). All p values <0.05 were considered significant, and all p values were 2-sided.

3. Results

3.1. Study participants and questionnaire acceptability

Overall, 308 completed questionnaires were analyzed out of 492 patients (63%). Median patient age was 55 years (interquartile range of 38 to 64 years old), with a minimum of 18 and a maximum of 89 years old; 161 (52.3%) were females, 197 (64%) were married, 233 (75.6%) had received higher education. Female doctor consulted 241 (78.2%) patients.

A total of 308 surveys were analyzed, and 45 item responses were missing. These represented 14.2% ('emotional' questions represented 6.8% of these) of the data to be missing. The result of Little's MCAR test was negative ($\chi 2 = 181.1$; $p = 0.017$), and since there was no specific pattern of missingness, we consumed that the data were missed at random (MAR, not 'completely' at random).

Acceptability was measured by the mean time needed to complete the PEQ and the response rate.

A typical survey response rate is usually from 5 to 30%, and a 50% response is excellent. As our questionnaire polling was stricter than a classic online survey, the expert panel decided that the cut-off of 50% response was acceptable for this study. As stated above, the response rate was 63%, deemed appropriate.

The authors of the original research did not provide a precise time to complete the questionnaire. They have just stated that "…[it] can be completed in a few minutes in the waiting room". We used the mean completion time from non-medical specialists (participants of the study's first phase). We assumed they would complete the questionnaire quicker than 'real' patients. Patients were expected to fill in the questionnaire no more than three times longer (i.e., 9 min). The mean time of a webpage session in the mass survey did not exceed 7 min (from 4 to 7 min).

3.2. Validity

As the data was skewed (-2.0 ± 0.98) and the kurtosis was also pronounced (4.79 ± 0.28), non-parametric methods were deemed appropriate. The results of Bartlett's test of sphericity indicated that the correlation matrix was not random, $\chi 2 = 2.648$, $p < 0.001$, and the KMO statistic was 0.80, well above the minimum standard for conducting factor analysis. Therefore, it was determined that the correlation matrix was appropriate for factor analysis.

As a result of the Parallel Analysis, four factors were extracted (Fig. 1).

3.2.1. Exploratory factor analysis

As the first step, we tried to build a simple structure with a promax (oblique) technique. Unfortunately, the results were inadequate: while the same items formed the first three domains, the fourth domain ('Communication barriers') comprised both useful and unsuitable items. The latter was a question from the 'Experience with the auxiliary staff'. Another question from this domain has had low factor loading. Moreover, the promax rotation has given us only a 59% explanation of the total variance.

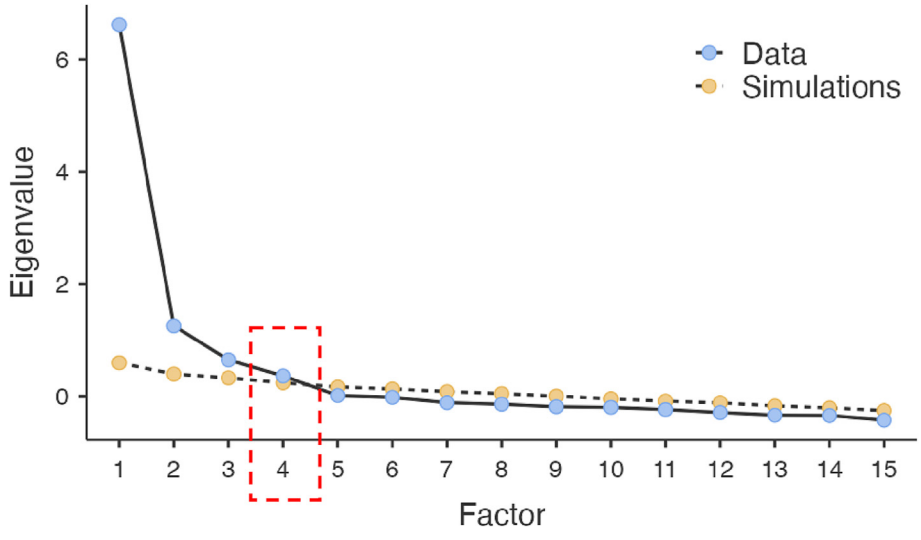


Fig. 1. Cattel's Scree Test.

On the other hand, we have built an adequate exploratory factor loading model when the varimax rotation was applied (Table 1 **and** Fig. 2). With the varimax rotation the model explained 65.4% of the total variance.

We revealed that the factor loadings of questions Q_3 (*Will you be able to handle your health problems differently?*), Q_13 (*I sensed that other patients could listen in when I was talking to the staff*), and Q_14 (*I felt like one of the crowd*) were constantly below the threshold of 0.3 or within 0.3–0.5. In addition, it was these questions that turned out to be attributed to different or even several factors at once at each iteration. In the original questionnaire, Q_3 should have been assigned to the factor 'Outcome of this specific visit', and questions Q_13 and Q_14 formed a separate factor 'Experience with the auxiliary staff'. After these questions were excluded, the remaining 15 were retained and formed a model. The first factor is mainly loaded by items Q_1, Q_2, and Q_4, which refer to the outcomes of this specific visit. The second factor is loaded with Q_5 to Q_8 as in the original PEQ and reflects communication experiences. The third factor encompasses items Q_9 to Q_12, again similar to the original questionnaire. However, the name of the factor was changed from 'Communication barriers' to 'Communication competency' as these questions sounded different, and this combination of questions denoted another term when translated into Russian. Finally, the fourth factor retained all 4 questions of the emotional sphere (Q_15-Q_18). Correlations between each item and the related domain indicated adequate discrimination for each item (> 0.5 for each of the retained questions).

3.2.2. Partial confirmatory factor analysis

We did not perform an additional survey to obtain data for classical confirmatory analysis. Our study was underpowered to randomly divide the cohort as fewer than 200 participants would be in either subgroup. Therefore, we decided to conduct a pCFA at the first stage of the work, as this is a feasible option [28].

As shown in Table 2, the standardized factor loadings for all the retained questions were >0.5. Moreover, most fit indexes met the requirements (Table 3).

The four-factor structure of the questionnaire remained stable without any dropouts. There was a fair-to-moderate but significant correlation between all four factors (inter-factor correlation <0.59 for each pair of the factors), indicating that the outcomes, communication, and emotions are closely related to each other (Fig. 3). It is worth noting that the mean inter-factor correlation was lower than within-factor correlation (0.718).

3.3. Reliability

The Cronbach's alpha and McDonald's omega coefficients of the overall scale were 0.912 and 0.921, indicating high internal consistency (Table 4). As shown in Table 5, there was no substantial improvement after each item's removal (coefficients between 0.910 and 0.920).

Table 1
Results of an exploratory factor analysis compared with original PEQ.

Question		Factor								Item-Factor correlation
		Outcome of this specific visit (F1)		Communication experiences (F2)		Communication competency (F3)		Emotions after this visit (F4)		
Q_1	Do you know what to do to reduce your health problem(s)?	0.79	0.82							0.78
Q_2	Do you know what to expect from now on?	0.92	0.82							0.83
Q_4	Will it [this specific visit] lead to fewer health problem(s)?	0.51	0.67							0.88
Q_5	We had a good talk [with the doctor]			0.78	0.83					0.62
Q_6	I felt reassured			0.73	0.84					0.85
Q_7	The doctor understood what was on my mind			0.81	0.78					0.80
Q_8	I felt I was taken care of			0.86	0.84					0.68
Q_9	It was a bit difficult to connect with the doctor					0.54	0.64			0.69
Q_10	Too much time was spent on small talk					0.54	0.82			0.62
Q_11	It was a bit difficult to ask questions					0.77	0.76			0.77
Q_12	Important decisions were made over my head					0.66	0.72			0.79
Q_15	After this visit I felt: relieved or worried							0.60	0.74	0.80
Q_16	After this visit I felt: sad or cheerful							0.87	0.82	0.88
Q_17	After this visit I felt: strenthened or worn out							0.74	0.78	0.86
Q_18	After this visit I felt: relaxed or tense							0.74	0.73	0.84

Footnote: "Minimum residual" extraction method was used in combination with an oblimin rotation; Q_n - number of a question. The factor loadings of the original PEQ are given on white background.
Questions 3, 13, 14 are excluded (see in text).

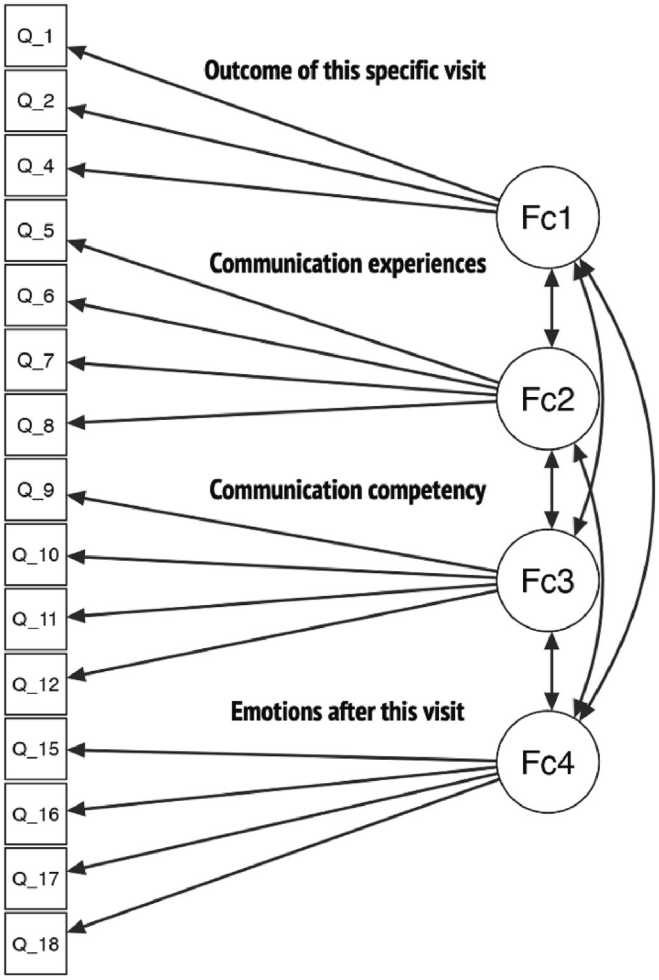


Fig. 2. Visual representation of the questions and their related factors according factor loadings (Path diagram). Footnotes: Q_n – number of the question; Fc_n – serial number of the factor with the brand-new Factor names.

Fifty-three percent of patients were satisfied with their care. The mean score for the outcome domain was 4.34 ± 0.95 (95% CI [4.24;4.45]), for communication experiences 4.74 ± 0.71 (95% CI [4.66;4.82]), for communication competency 4.63 ± 0.77 (95% CI [4.55;4.72]) and for emotional domain 3.88 ± 0.96 (95% CI [3.77;3.98]). Merely 16.2% of patients described their experiences with communication as less than optimal. Up to 16% of patients reported communication incompetency, and 42% had not experienced the optimal outcome after the encounter. Up to 35.7% of patients left the Center with negative or no positive feelings. Just 2.6% scored below three on the total scale. No differences were noted between the groups analyzed (according to sex, doctor's gender, education, and marital status) neither when the mean scoring was applied nor with the total score of the questionnaire.

Table 2
Partial confirmatory factor analysis.

Factor	Question	Estimate	SE	Z	*p*-value
Factor 1	Q_1	0.799	0.0422	18.9	<0.001
Outcome of this specific visit	Q_2	0.932	0.0504	18.5	<0.001
	Q_4	0.762	0.0549	13.9	<0.001
Factor 2	Q_5	0.626	0.0348	18.0	<0.001
Communication experiences	Q_6	0.725	0.0393	18.4	<0.001
	Q_7	0.697	0.0376	18.5	<0.001
	Q_8	0.683	0.0364	18.7	<0.001
Factor 3	Q_9	0.710	0.0505	14.1	<0.001
Communication competency	Q_10	0.594	0.0416	14.3	<0.001
	Q_11	0.689	0.0458	15.0	<0.001
	Q_12	0.743	0.0613	12.1	<0.001
Factor 4	Q_15	0.879	0.0648	13.6	<0.001
Emotions after this visit	Q_16	0.906	0.0569	15.9	<0.001
	Q_17	0.867	0.0545	15.9	<0.001
	Q_18	0.882	0.0581	15.2	<0.001

Footnote: Q_n - number of a question; SE - standard error of a mean; *Z*- value of the Fisher's test.

Table 3
Goodness-of-fit statistic and the fit indexes.

CFI	TLI	SRMR	RMSEA	RMSEA 90% CI		Test for exact fit		
				Lower	Upper	χ^2	df	p-value
0.94	**0.92**	**0.048**	**0.083**	**0.072**	**0.094**	264	84	**<0.001**

Footnotes: CFI - Comparative fit index; TLI - Tucker-Lewis index (Non-normed fit index); SRMR - standardized root mean square residual; RMSEA - root mean square error of approximation (with 90% confidence interval (CI); df - degrees of freedom.

The Russian translation of the PEQ can be found in the **Appendix** section. The scoring instructions are simple. The overall score can be calculated by quickly summing all the points item-by-item or dividing items by the number of items (i.e., the mean score).

4. Discussion and conclusion

4.1. Discussion

Approaches to assessing the quality of care for outpatients in the Russian Federation still need to be fully formed, and therefore the development of such instruments is of particular interest. In this study, we have consistently passed the mandatory initial stage of cultural adaptation and internal validation of the PREM for outpatients. The cross-cultural adaptation process included expert team and patient discussions, a pilot testing, thus ensuring the conceptual and linguistic equivalence of the translated questionnaire. The face and content validity of the original tool was confirmed both in pilot and general surveys. In the qualitative phase of the study, the questionnaire was restructured. During the quantitative phase, some questions were deemed redundant or irrelevant to Russian-language patients. These first assumptions should now be externally validated in the more extensive and differing cohorts of patients.

The choice of the questionnaire is quite understandable and arises from the two rationales: theoretical and practical.

The theoretical rationale is that the PEQ was one of the first PREMs of its kind, which was not limited to determining only satisfaction with the treatment rather than focusing on the whole process of doctor encounters. Moreover, the questionnaire was designed '…in line with patient's preferred content and using their own words and statements', which is relevant to the VBM approach. The original paper's authors shifted their vision from capturing satisfaction to more personal, affective responses. Moreover, satisfaction with care is mainly driven by social preconceptions and consumerism. Since medical consumerism is not very prominent in today's Russia, but social preconceptions prevail, the rationale from the PEQ authors seems adequate. Like in Norway in the 2000s, healthcare in the Russian Federation is not as competitive as in the USA or Europe, and it is primarily true for large cities. Therefore, like the PEQ authors, we believed we did not need another satisfaction questionnaire. Theory is supported by a proper statistical analysis. In the most recent systematic review Derriennic J et al. provide an overview of 29 questionnaires concerning quality of medical care [29]. Although, most of questionnaires are satisfy ones from 1990s, it seems to us that the PEQ stands out by excellent structural validity, low responder-burden (only 18 questions in the original version).

The practical rationale for using the PEQ was its ease of use, which is crucial for the busy schedule in most ambulatory centers and clinics in the Russian Federation. The less time it takes to fill in, the fewer answers one can get, and the less is missing data. It is particularly true for our

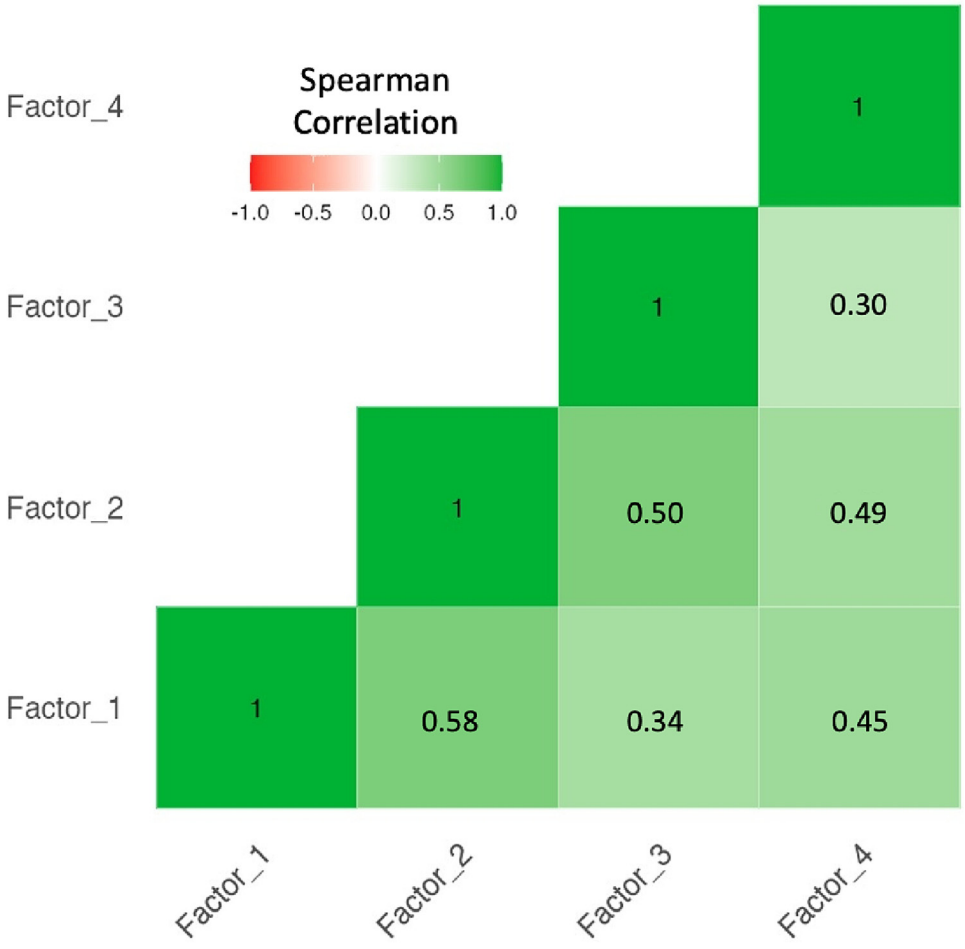


Fig. 3. Between factor correlation heatmap.

center: while patients give little attention to a substantial satisfactory survey, patients more readily fill in the short PEQ.

Different PREMs have been developed across the globe. Some researchers have mentioned constraints of PEQ and similar EUROPEP [30] that concerned their reliability and construct validity, respectively [7]. However, these issues were mentioned only because the authors wanted to develop a nationwide survey for which these tools were unsuitable. There are several such national surveys have been recently developed: for example, the Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) survey, which consists of 27 questions and is administered to a sample of hospital inpatients chosen at random, was created in 2002 by the Centers for Medicare and Medicaid Services and the Agency for Healthcare Research and Quality. With the Clinician and Group CAHPS (CG-CAHPS) survey, a version of the HCAHPS created for outpatient contacts, the emphasis on tracking and improving the patient experience has gradually moved to the outpatient setting [31]. These questionnaires, like NHS Patient Experience or Press Ganey Patient Satisfaction surveys, are complex and comprise dozens of questions that may seem irrelevant to patients [32], although their wide adoption led to improvement in patients' experiences with care [33].

On the contrary, disease-specific tools do not represent the quality of primary care for most patients; neither are they developed properly nor feasible [34]. We did not aim to develop either a national or disease-specific survey. We have decided to test the primary care questionnaire as the first step towards a better understanding of what matters for Russian ambulatory patients because, in Russia, ambulatory care is, like in most countries, the cornerstone of the whole healthcare system.

The short PEQ is easy to complete. Given the average time required to fill in the questionnaire (~7 min), it can be provided to patients immediately after the visit to the facility or, as we found out, using a simple and widely available online option. Online polling is one of the strengths of our study. As many patient-experience surveys are conducted by telephone, thus adding more personnel and decreasing the reliability of the proxy-derived data [35], online-survey may serve as an alternative to obtain more reliable data with a higher response rate [36,37].

Some questions were excluded from the original PEQ questionnaire. We eliminated the fifth domain (Experience with the auxiliary staff) after

Table 4
Item reliability.

		If the item is dropped	
Question	Item-factor correlation	Cronbach's α	McDonald's ω
Q_1	0.766	0.904	0.914
Q_2	0.785	0.905	0.915
Q_4	0.654	0.905	0.915
Q_5	0.797	0.905	0.912
Q_6	0.802	0.902	0.910
Q_7	0.818	0.903	0.911
Q_8	0.824	0.903	0.911
Q_9	0.621	0.908	0.917
Q_10	0.635	0.908	0.917
Q_11	0.699	0.907	0.916
Q_12	0.595	0.911	0.919
Q_15	0.636	0.909	0.918
Q_16	0.735	0.910	0.920
Q_17	0.722	0.907	0.918
Q_18	0.688	0.907	0.918

Footnote: Q_n - number of a question.

Table 5
Internal consistency compared with the original PEQ.

	Cronbach's α		McDonald's ω
Whole scale	**0.912**	**0.82**	**0.921**
Outcome of this specific visit (F1)	0.856	0.80	0.866
Communication experiences (F2)	0.917	0.9	0.918
Communication competency (F3)	0.810	0.77	0.820
Emotions after this visit (F4)	0.852	0.82	0.856

Footnotes: The indicators of internal consistency of the original PEQ are given on white background.

retracting Q_13 (*I sensed that other patients could listen in when I was talking to the staff*) and Q_14 (*I felt like one of the crowd*). The peculiarities of outpatient care in Russia can explain this exclusion. Almost all medical procedures, whether the history taking or physical examination, are carried out by a doctor, not by other medical or paramedical staff (i.e., a nurse or nursing assistant). The whole appointment takes place in a separate room. Occasionally (more often in public clinics called 'polyclinics'), a nurse is present in the doctor's office to fill in the information into the medical information system. Therefore, relevant questions for patients in Europe or the USA turned out to be obscure to Russian outpatients. On the other hand, these items might be engaging to in-patients because the situation in hospitals is the opposite. There are usually 3 or 4 patients in the ward in Russian hospitals, so they may feel uncomfortable talking to their doctor in front of others. Although when using promax rotation, the factor loadings of variables were higher than with varimax, one of the 'Auxiliary staff' questions (Q_13) was falsely assigned to one of the retained factors (in the original PEQ, this factor was called 'Communication barriers') and another one (Q_14) had a low factor loading, and thus excluded from the model. In our opinion, this model did not make much sense, as the question regarding comfortability and privacy is unrelated to the 'barrier' when it comes to a communication with the doctor. Along with low cumulative variance, the promax rotation was deemed inappropriate. While some researchers state that oblique rotation is highly accurate, others point out that orthogonal rotation produces more easily interpretable results and is slightly simpler [38]. An additional explanation of favoring varimax rotation is that the authors of the original research used this method during validation process, however, without clear reasoning. Even assuming the presets of varimax rotation, all the factors correlated with each other but less than the items of which they were built.

Another question (Q_3 - *Will you be able to handle your health problems differently?)* was excluded from the first domain because of the high skewness and low factor loading. When analyzing the statements of the Q_1 (*Do you know what to do to reduce your health problem(s)?*) and the Q_3 (*Will you be able to handle your health problems differently?*) sounded and formally meant similar in Russian. The name of the fourth domain, called *Communication barriers*, was replaced by *Communication competency*. If it is difficult to establish communication to ask questions, and the doctor talks a lot about unimportant matters while simultaneously making treatment decisions on his own, it is likely because of the doctor's communicative incompetence, not of the barrier itself. Norwegian PEQ was also taken as the doctor's technical skills are irrelevant for Russian patients because they take doctors' qualifications for granted, and communicative competency comes to the fore.

It is intriguing that patients in our study were mostly satisfied with the outcomes of the appointment and experienced good communication with doctors. These results contrast with the original research, where almost half of the patients needed more understanding of what to do and anticipate concerning their medical issues. They did not believe the visit would lead to fewer troubles or that it would significantly change their health behaviors. These may be due to the selection bias (large reference center with highly qualified staff) and because the practice of medicine has changed in the last 20 years (the original PEQ was developed in 2001). In our research almost all the domains have the same mean scores as in the original analysis except the Communication barriers (or Communication competency in our variant). The emotional domain had fewer answers on the extremes (30% scored 5, and 9.4% scored 1 or 2). It may be because patients' illnesses were mostly longstanding, and they did not expect surprising news about their health state and treatment, so 60% scored 3 or 4 instead.

In the original PEQ questionnaire [15], the authors faced the problem of pronounced skewness of questions in the emotional scale and decided to replace the 5-point with a 7-point rating scale. After this, factor loadings were acceptable for half of the initially included items (5 out of 10) and decreased the percentage of missing data. We have done the opposite, and it was the right thing to do. Only in the first question of this domain (*After this visit, I felt: relieved or worried*) was high negative skewness observed, with 55.8% of patients being mostly relieved. This result may occur since our center is not a reference for oncology outpatients. Patients attending our clinic are mostly consulted for other non-communicable diseases with a better prognosis. Although the rest of the questions followed the non-parametric distribution, they were not so highly skewed. The mean skewness and kurtosis and their standard deviations fall in acceptable the cut-offs for performing structural equation modelling [39].

We found no differences between pre-specified patient groups (according to patient gender, physician gender, marital status, and educational level). It limited the assessment of responsiveness and construct validity. Examining hypothesized differences in scores by additional patient characteristics (for example, the reason for the encounter, prior medical experiences, and self-assessed health-related quality of life) and by characteristics of the medical visit will provide additional evidence. The questionnaire's ability to identify variations between specific medical professionals (within one medical facility), medical units, or even clinics is equally crucial.

4.1.1. Limitations

To correctly interpret the study's results, it is necessary to identify its possible limitations. Despite the sufficient sample size, it was impossible to avoid selection bias and limited generalizability since the patient enrollment and surveying occurred in a metropolitan city and at a large outpatient center. Selection bias is inevitable in single-center studies. On the other hand, a doctor's appointment in our Centre is usually the first for most visiting patients, and many specialists consult patients on different health problems. Both provide patient diversity. To overcome selection bias, we encourage researchers to conduct further studies in different settings for external validation and to find possible discrepancies within different populations, thus providing evidence for future benchmarking. This idea is supported by the studies from Sweden, where the authors have found covariation between registered process measures and PREMs. It helped compare performance among providers as those highly ranked with process measures tend to be ranked high with PREMs (patient views correlated with clinical quality) [40].

Another prominent area for improvement is data and non-response bias. However, the response rate in our study was relatively high. The recent

meta-analysis of online surveys stated that the average response rate is about 44% [41], and we were close to this estimate. Data loss is unavoidable, and although primarily seen in surveys and retrospective studies, this is also a problem in modern randomized controlled trials with thousands of participants [42]. We have found that 14% of data were missed at random, half of them in the emotional sphere. We used the multiple imputation with regression method to correctly interpret the data. Studies of imputation methods with simulated and real data demonstrate that any method is probably effective when <5% of the data are missing, mean imputation is acceptable when <10% are missing, and regression imputation is acceptable when <15% of the data are missing. Multiple imputation is an effective method for dealing with missing data problems, according to excellent parameter estimation, variance estimation, and increased power. Multiple imputation is an effective method for dealing with missing data, according to excellent parameter estimation, variance estimation, and increased power. While some researchers say that 10% of missing data is the upper threshold [43], others claim that when 20% of data is missed at random, the difference between the bias of multiple imputation and the bias of complete case analysis is only 0.43 [44]. There are studies in which authors claim multiple imputation is unbiased even until 50% of the data are missed [45].

We did not ask patients validation questions like the authors of the original research (difference in simple satisfaction questions followed by analyzing the differences in all the five domains). Though it may be a surrogate measure of responsiveness, it would be more appropriate in our future research.

Finally, we used the English-language version of the PEQ instead of the Norwegian language. There are a lot of guidelines and reviews on translation-adaptation processes published to date. However, neither provides strong evidence of using the original (or so-called “mother”) language as the source for translation. As there is a trend towards globalization and harmonization of PRO and PRE tools, more and more researchers provide English versions of their questionnaires along with native versions. Although the Norwegian language was used in the original research, the English version was also available in a published paper [15]. The additional reasonings for English-to-Russian instead of Norwegian-to-Russian translation are the unequivocal sounding of the statements in the questionnaire and the lack of Norwegian native speakers in our research group. Furthermore, the result of the present work shows a fair item and conceptual equivalence and good psychometric properties of the adapted questionnaire. We are planning the next step of this research to confirm our preliminary results, and if in doubt, we will consider using the Norwegian version.

4.2. Innovation

The broader and more active implementation of tools for patient-perceived usefulness and healthcare delivery experience helps better understand the problems not detected by internal and external audits. Easy-to-fill questionnaires may stimulate the transition of VBM in the Russian Federation. From our perspective, any VBM-related reorganizational types like accountable care organizations, pay-for-performance (P4P), or bundled payment models should also include the patient's experience with care. Healthcare professionals should be able to develop and effectively implement solid strategies that place the patient at the center of health services delivery. While the core element of VBM is cost-effectiveness analysis which reaches mainly an incremental cost-utility ratio, it might be of particular interest to incorporate calculation of the incremental cost-effectiveness ratio based on ΔPREMs. As more and more private centers are opening in big cities, the standardization of patient experience and an established mechanism of comparing the clinics will inform potential clients (patients [*sic*]).

To our knowledge, this is the first attempt to translate PREM to Russian language which assesses the experience of patients with outpatient care. The area of PRE is unexplored in the Russian healthcare. We sought to perform a guideline-directed statistical analysis with a robust theoretical basis. So, our psychometric testing might be considered as the model for the future studies in this area of research. We have used emerging information and communication technology to enroll patients effectively and to get feedback from them (e.g., QR codes). Recent technological developments may enable PREMs to play an essential part in the future transformation of healthcare [46]. Our results suggest that even in our large and expert center, many patients receive medical care, which could be of higherquality as per their answers. One of the possible future objectives of the research might be an attempt to conduct structured interviews with patients who either scored low or high. Such a study will be helpful for doctor-patient training programs or continuous medical education.

Our research will be continued to confirm the initially developed model. Pilot results will form the basis for the next step – external validation in a wide range of hospitals and healthcare centers. So, more and more staff of the high-quality centers and smaller clinics across the Russian Federation will become familiar with the concept of PREM. The wider the second external validation part of the study, the more interest it will arouse, and more research in this area might be in the nearest future. Finally, the higher quality PREMs are developed or adapted, the higher the healthcare quality.

5. Conclusion

The content of the PEQ questionnaire was analyzed correctly and adjusted, tailoring the specifics of outpatient care in the Russian Federation. We performed initial steps in validation of the questionnaire. We are encouraged to further externally validate the PEQ for its proper use in assessing the patient's experience with a consultation service.

Funding details

This work was financially supported by the Ministry of Science and Higher Education of the Russian Federation (Agreement № 075-15-2022-301). The funding source was not involved in study design; in the collection, analysis, and interpretation of data; in the writing of the manuscript; and in the decision to submit the article for publication.

Consent for publication

All co-authors have given their consent for this paper to be published.

Informed Consent and patient details

We, the authors, confirm verbal and/or electronic informed consents was obtained in every patient before any study procedure was performed. We also confirm all patient/personal identifiers have been removed or disguised so the patient/person(s) described are not identifiable and cannot be identified through the details of the story.

CRediT authorship contribution statement

Mikhail Ionov: Conceptualization, Methodology, Investigation, Visualization, Writing – original draft. **Elena Dubinina:** Methodology, Validation, Supervision. **Ilya Tregubenko:** Formal analysis, Validation, Supervision. **Nadezhda Zvartau:** Writing – review & editing, Project administration. **Alexandra Konradi:** Funding acquisition, Project administration.

Declaration of Competing Interest

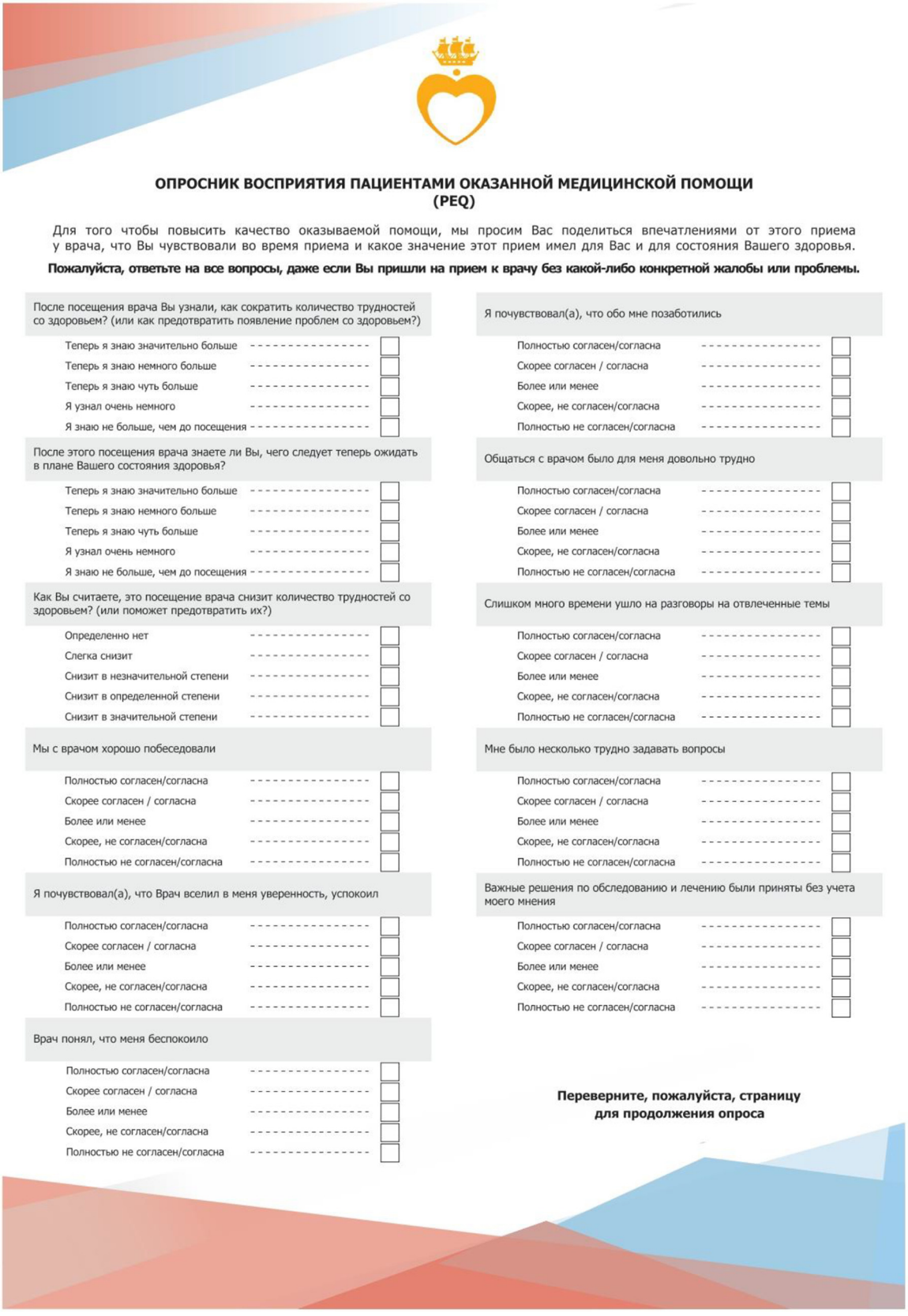
The authors hereby declare that no competing interests exist.

Acknowledgements

The authors would like to thank all the patients-participants. The authors are also grateful to Dr. Dmitry Ovchinnikov (the Head of Clinical and Diagnostic Centre) and to Mr. Nicolay Smirnov (IT-specialist) for their contribution in patient's recruitment and in data mining, respectively.

Appendix

The Russian translation of PEQ questionnaire



ОПРОСНИК ВОСПРИЯТИЯ ПАЦИЕНТАМИ ОКАЗАННОЙ МЕДИЦИНСКОЙ ПОМОЩИ (PEQ)

Для того чтобы повысить качество оказываемой помощи, мы просим Вас поделиться впечатлениями от этого приема у врача, что Вы чувствовали во время приема и какое значение этот прием имел для Вас и для состояния Вашего здоровья.

Пожалуйста, ответьте на все вопросы, даже если Вы пришли на прием к врачу без какой-либо конкретной жалобы или проблемы.

После посещения врача Вы узнали, как сократить количество трудностей со здоровьем? (или как предотвратить появление проблем со здоровьем?)

- Теперь я знаю значительно больше ☐
- Теперь я знаю немного больше ☐
- Теперь я знаю чуть больше ☐
- Я узнал очень немного ☐
- Я знаю не больше, чем до посещения ☐

После этого посещения врача знаете ли Вы, чего следует теперь ожидать в плане Вашего состояния здоровья?

- Теперь я знаю значительно больше ☐
- Теперь я знаю немного больше ☐
- Теперь я знаю чуть больше ☐
- Я узнал очень немного ☐
- Я знаю не больше, чем до посещения ☐

Как Вы считаете, это посещение врача снизит количество трудностей со здоровьем? (или поможет предотвратить их?)

- Определенно нет ☐
- Слегка снизит ☐
- Снизит в незначительной степени ☐
- Снизит в определенной степени ☐
- Снизит в значительной степени ☐

Мы с врачом хорошо побеседовали

- Полностью согласен/согласна ☐
- Скорее согласен / согласна ☐
- Более или менее ☐
- Скорее, не согласен/согласна ☐
- Полностью не согласен/согласна ☐

Я почувствовал(а), что Врач вселил в меня уверенность, успокоил

- Полностью согласен/согласна ☐
- Скорее согласен / согласна ☐
- Более или менее ☐
- Скорее, не согласен/согласна ☐
- Полностью не согласен/согласна ☐

Врач понял, что меня беспокоило

- Полностью согласен/согласна ☐
- Скорее согласен / согласна ☐
- Более или менее ☐
- Скорее, не согласен/согласна ☐
- Полностью не согласен/согласна ☐

Я почувствовал(а), что обо мне позаботились

- Полностью согласен/согласна ☐
- Скорее согласен / согласна ☐
- Более или менее ☐
- Скорее, не согласен/согласна ☐
- Полностью не согласен/согласна ☐

Общаться с врачом было для меня довольно трудно

- Полностью согласен/согласна ☐
- Скорее согласен / согласна ☐
- Более или менее ☐
- Скорее, не согласен/согласна ☐
- Полностью не согласен/согласна ☐

Слишком много времени ушло на разговоры на отвлеченные темы

- Полностью согласен/согласна ☐
- Скорее согласен / согласна ☐
- Более или менее ☐
- Скорее, не согласен/согласна ☐
- Полностью не согласен/согласна ☐

Мне было несколько трудно задавать вопросы

- Полностью согласен/согласна ☐
- Скорее согласен / согласна ☐
- Более или менее ☐
- Скорее, не согласен/согласна ☐
- Полностью не согласен/согласна ☐

Важные решения по обследованию и лечению были приняты без учета моего мнения

- Полностью согласен/согласна ☐
- Скорее согласен / согласна ☐
- Более или менее ☐
- Скорее, не согласен/согласна ☐
- Полностью не согласен/согласна ☐

Переверните, пожалуйста, страницу для продолжения опроса

The Russian-language version of the Norwegian "Patient Experience Questionnaire"

References

[1] Porter ME. What is value in health care? N Engl J Med. 2010;363:2477–81. https://doi.org/10.1056/NEJMp1011024.

[2] Coulter A. Measuring what matters to patients. BMJ. 2017;356:j816. https://doi.org/10.1136/bmj.j816.

[3] Sixma HJ, Kerssens JJ, van Campen C, Peters L. Quality of care from the patients' perspective: from theoretical concept to a new measuring instrument. Health Expect. 1998;1:82–95. https://doi.org/10.1046/j.1369-6513.1998.00004.x.

[4] Kingsley C, Patel S. Patient-reported outcome measures and patient-reported experience measures; 2017.https://doi.org/10.1093/bjaed/mkw060.

[5] Bele S, Teela L, Zhang M, Rabi S, Ahmed S, van Oers HA, et al. Use of patient-reported experience measures in pediatric care: a systematic review. Front Pediatr. 2021;9. https://doi.org/10.3389/fped.2021.753536. (accessed October 12, 2022).

[6] Anhang Price R, Elliott MN, Zaslavsky AM, Hays RD, Lehrman WG, Rybowski L, et al. Examining the role of patient experience surveys in measuring health care quality. Med Care Res Rev. 2014;71:522–54. https://doi.org/10.1177/1077558714541480.

[7] Holmboe O, Iversen HH, Danielsen K, Bjertnaes O. The Norwegian patient experiences with GP questionnaire (PEQ-GP): reliability and construct validity following a national survey. BMJ Open. 2017;7:e016644. https://doi.org/10.1136/bmjopen-2017-016644.

[8] 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. https://doi.org/10.1136/bmjopen-2012-001570.

[9] Black N, Varaganum M, Hutchings A. Relationship between patient reported experience (PREMs) and patient reported outcomes (PROMs) in elective surgery. BMJ Qual Saf. 2014;23:534–42. https://doi.org/10.1136/bmjqs-2013-002707.

[10] Bull C, Byrnes J, Hettiarachchi R, Downes M. A systematic review of the validity and reliability of patient-reported experience measures. Health Serv Res. 2019;54:1023–35. https://doi.org/10.1111/1475-6773.13187.

[11] Federal law no.323-FZ "On Fundumentals of Health Protection in the Russian Federation" dated November 21. https://minzdrav.gov.ru/documents/7025-federalnyy-zakon-323-fz-ot-21-noyabrya-2011-g; 2011. [accessed April 15, 2023].

[12] Questionnaire for assessing the quality of conditions for the provision of services by medical organizations in an outpatient setting. https://anketa.minzdrav.gov.ru/staticogvjustank/64/1; 2023. (accessed April 15, 2023).

[13] Beattie M, Murphy DJ, Atherton I, Lauder W. Instruments to measure patient experience of healthcare quality in hospitals: a systematic review. Syst Rev. 2015;4:97. https://doi.org/10.1186/s13643-015-0089-0.

[14] Gjersing L, Caplehorn JRM, Clausen T. Cross-cultural adaptation of research instruments: language, setting, time and statistical considerations. BMC Med Res Methodol. 2010;10:13. https://doi.org/10.1186/1471-2288-10-13.

[15] Steine S, Finset A, Laerum E. 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. https://doi.org/10.1093/fampra/18.4.410.

[16] Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, et al. ISPOR task force for translation and cultural adaptation, principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR task force for translation and cultural adaptation. Value Health. 2005;8:94–104. https://doi.org/10.1111/j.1524-4733.2005.04054.x.

[17] Baker M. In other words: A coursebook on translation. 3rd ed. London: Routledge; 2018. https://doi.org/10.4324/9781315619187.

[18] Pedersen AB, Mikkelsen EM, Cronin-Fenton D, Kristensen NR, Pham TM, Pedersen L, et al. Missing data and multiple imputation in clinical epidemiological research. Clin Epidemiol. 2017;9:157–66. https://doi.org/10.2147/CLEP.S129785.

[19] Harrell FE. Missing Data. In: Harrell FE, editor. Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis. New York, NY: Springer; 2001. p. 41–52. https://doi.org/10.1007/978-1-4757-3462-1_3.

[20] Schumacker RE. Learning Statistics Using R. 1 Oliver's Yard, 55 City Road London EC1Y 1SP: SAGE Publications, Inc; 2015.https://doi.org/10.4135/9781506300160.

[21] Bartlett MS. A note on the multiplying factors for various $\chi 2$ approximations. J R Stat Soc B Methodol. 1954;16:296–8. https://doi.org/10.1111/j.2517-6161.1954.tb00174.x.

[22] Kaiser HF. The varimax criterion for analytic rotation in factor analysis. Psychometrika. 1958;23:187–200. https://doi.org/10.1007/BF02289233.

[23] Horn JL. A rationale and test for the number of factors in factor analysis. Psychometrika. 1965;30:179–85. https://doi.org/10.1007/BF02289447.

[24] Cattell RB. The scree test for the number of factors. Multivar Behav Res. 1966;1:245–76. https://doi.org/10.1207/s15327906mbr0102_10.

[25] Norman GR, Streiner DL, Norman GR. Biostatistics: The bare essentials. http://site.ebrary.com/id/10949549; 2014. (accessed April 14, 2022).

[26] Watkins MW. Exploratory factor analysis: a guide to best practice. J Black Psychol. 2018;44:219–46. https://doi.org/10.1177/0095798418771807.

[27] Fabrigar LR, Wegener DT, MacCallum RC, Strahan EJ. Evaluating the use of exploratory factor analysis in psychological research. Psychol Methods. 1999;4:272–99. https://doi.org/10.1037/1082-989X.4.3.272.

[28] Gignac GE. Partial confirmatory factor analysis: described and illustrated on the NEO-PI-R. J Pers Assess. 2009;91:40–7. https://doi.org/10.1080/00223890802484126.

[29] Derriennic J, Nabbe P, Barais M, Le Goff D, Pourtau T, Penpennic B, et al. A systematic literature review of patient self-assessment instruments concerning quality of primary care in multiprofessional clinics. Fam Pract. 2022;39:951–63. https://doi.org/10.1093/fampra/cmac007.

[30] Wensing M, Mainz J, Grol R. A standardised instrument for patient evaluations of general practice care in Europe. Eur J Gen Pract. 2000;6:82–7. https://doi.org/10.3109/13814780009069953.

[31] CAHPS Clinician & Group Survey. Internet source. https://www.ahrq.gov/cahps/surveys-guidance/cg/index.html; 2023. (accessed April 15, 2023).

[32] Boiko O, Campbell JL, Elmore N, Davey AF, Roland M, Burt J. The role of patient experience surveys in quality assurance and improvement: a focus group study in English general practice. Health Expect. 2015;18:1982–94. https://doi.org/10.1111/hex.12298.

[33] Elliott MN, Lehrman WG, Goldstein EH, Giordano LA, Beckett MK, Cohea CW, et al. Hospital survey shows improvements in patient experience. Health Aff (Millwood). 2010;29:2061–7. https://doi.org/10.1377/hlthaff.2009.0876.

[34] Cornelis C, den Hartog SJ, Bastemeijer CM, Roozenbeek B, Nederkoorn PJ, Van den Berg-Vos RM. Patient-reported experience measures in stroke care. Stroke. 2021;52:2432–5. https://doi.org/10.1161/STROKEAHA.120.034028.

[35] Li M, Harris I, Lu ZK. Differences in proxy-reported and patient-reported outcomes: assessing health and functional status among medicare beneficiaries. BMC Med Res Methodol. 2015;15:62. https://doi.org/10.1186/s12874-015-0053-7.

[36] Labovitz J, Patel N, Santander I. Web-based patient experience surveys to enhance response rates a prospective study. J Am Podiatr Med Assoc. 2017;107:516–21. https://doi.org/10.7547/16-001.

[37] Lin Y, Hong YA, Henson BS, Stevenson RD, Hong S, Lyu T, et al. Assessing patient experience and healthcare quality of dental care using patient online reviews in the United States: mixed methods study. J Med Internet Res. 2020;22:e18652. https://doi.org/10.2196/18652.

[38] Taherdoost H, Sahibuddin S, Jalaliyoon N. Exploratory factor analysis; concepts and theory (august 1, 2022). Adv Appl Pure Math. 2022;27:375–82. Available at SSRN:. https://ssrn.com/abstract=4178683.

[39] Brown TA. Confirmatory factor analysis for applied research. New York, NY, US: The Guilford Press; 2006.

[40] Glenngård AH, Anell A. Process measures or patient reported experience measures (PREMs) for comparing performance across providers? A study of measures related to access and continuity in Swedish primary care. Prim Health Care Res Dev. 2018;19:23–32. https://doi.org/10.1017/S1463423617000457.

[41] Wu M-J, Zhao K, Fils-Aime F. Response rates of online surveys in published research: a meta-analysis. Comput Hum Behav Rep. 2022;7:100206. https://doi.org/10.1016/j.chbr.2022.100206.

[42] Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ. 2009;338:b2393. https://doi.org/10.1136/bmj.b2393.

[43] Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomised clinical trials - a practical guide with flowcharts. BMC Med Res Methodol. 2017;17:162. https://doi.org/10.1186/s12874-017-0442-1.

[44] Lee JH, Huber JC. Evaluation of multiple imputation with large proportions of missing data: how much is too much? Iran J Public Health. 2021;50:1372–80. https://doi.org/10.18502/ijph.v50i7.6626.

[45] De Silva AP, Moreno-Betancur M, De Livera AM, Lee KJ, Simpson JA. A comparison of multiple imputation methods for handling missing values in longitudinal data in the presence of a time-varying covariate with a non-linear association with time: a simulation study. BMC Med Res Methodol. 2017;17:114. https://doi.org/10.1186/s12874-017-0372-y.

[46] Teisberg E, Wallace S, O'Hara S. Defining and implementing value-based health care: a strategic framework. Acad Med. 2020;95:682–5. https://doi.org/10.1097/ACM.0000000000003122.