Psychometric Testing of the Primary Care Outcomes Questionnaire

Authors

<table>
<thead>
<tr>
<th>Name</th>
<th>Advanced Degree</th>
<th>Institutional Affiliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mairead Murphy</td>
<td>PhD</td>
<td>University of Bristol</td>
</tr>
<tr>
<td>Sandra Hollinghurst</td>
<td>PhD</td>
<td>University of Bristol</td>
</tr>
<tr>
<td>Sean Cowlishaw</td>
<td>PhD</td>
<td>University of Bristol</td>
</tr>
<tr>
<td>Chris Salisbury</td>
<td>MSc, MD, FRCGP</td>
<td>University of Bristol</td>
</tr>
</tbody>
</table>

Abstract

Background

Patients attend primary care for many reasons and to achieve a range of possible outcomes. There is currently no patient-reported outcome measure designed to capture these diverse outcomes, and trials of interventions in primary care may thus fail to detect beneficial effects.

Aim

This study describes the psychometric testing of the Primary Care Outcomes Questionnaire (PCOQ), which was designed to capture a broad range of outcomes relevant to primary care.

Design and Setting

Questionnaires administered in primary care in South West England.

Methods

Patients completed the PCOQ in GP waiting rooms before a consultation, and a second questionnaire including the PCOQ and seven comparator PROMs after one week. Psychometric testing included exploratory factor analysis on the PCOQ, internal consistency, correlation coefficients between domain scores and comparator measures, and repeated measures effect-sizes indicating change across one week.
Results

602 patients completed the PCOQ at baseline, and 264 (44%) returned the follow-up questionnaire. Exploratory factor analysis suggested four dimensions underlying the PCOQ items: Health and Well-Being, Health Knowledge and Self-Care, Confidence in Health Provision, and Confidence in Health Plan. Each dimension was internally consistent and correlated as expected with comparator PROMs, providing evidence of construct validity. Patients reporting an improvement in their main problem exhibited small to moderate improvements in relevant domain scores on the PCOQ.

Conclusion

The PCOQ was acceptable, feasible, showed strong psychometric properties and was responsive to change. It is a promising new tool for assessment of outcomes of primary care interventions from a patient perspective.

Keywords

Primary care, patient-centred care, Psychometrics, Questionnaires, Patient-reported outcomes, Health care delivery/HSR: quality of care

How this fits in

Patients attend primary care with many types of problems and to achieve a range of possible outcomes, but there is currently no patient-reported outcome measure designed to capture these diverse outcomes. As such, trials of interventions in primary care may fail to detect beneficial effects. The PCOQ was developed to measure a range of outcomes commonly influenced in primary care including health and well-being, health knowledge and self-care, confidence in health provision and confidence in health plan. Testing showed it was acceptable, feasible and had strong psychometric properties including responsiveness to change. It is a promising new tool for assessment of outcomes of primary care interventions from a patient perspective.
Primary care has been evolving in recent years to meet changing population needs and public expectations.\(^{1-6}\) As health services globally contend with aging populations and increasing multimorbidity,\(^{7}\) there have been sustained endeavours to improve service quality, costs, and outcomes in primary care.\(^{1}\) Innovations include electronic consultations,\(^{8}\) health coaching and behavioural change therapies,\(^{9}\) and interventions addressing frequent attenders.\(^{10}\)

Assessing the effectiveness of such interventions from a patient perspective involves patient-reported outcome measures (PROMs). Many PROMs are disease-specific and tailored to the symptoms and impacts on function of a particular condition.\(^{11}\) These are of limited value in studies where patients have various conditions. As a first contact, comprehensive and co-ordinating service,\(^{12}\) primary care requires a generic PROM which can be administered regardless of condition. This should be suitable for large-scale trials, based on outcomes that matter to patients and are influenced by GPs. It should also be “responsive”: i.e. able to detect changes over time.\(^{13}\)

A problem with many generic PROMs is that they are limited to symptoms and function. Primary care patients frequently present with other problems\(^{14}\) and many have long-term conditions,\(^{7,15,16}\) whereby improvement in function may be unrealistic. Leading generic PROMs such as the SF-36\(^{17}\) and EQ-5D\(^{18}\), therefore often show no change following interventions in primary care.\(^{19-21}\) Because of this, alternative measures have been designed specifically for primary care.\(^{21-23}\)

For example, the Measure Yourself Medical Outcome Profile (MYMOP) allows patients to specify their problems and shows change when other PROMs do not.\(^{21}\) However, MYMOP is administered through interviews, making it unfeasible for trials. It also remains limited to symptoms and function. In contrast, the Patient Enablement Measure (PEI) encompasses broader outcomes related to coping, understanding and confidence in health (but does not capture symptoms or function). Although it has been validated for primary care,\(^{22}\) the PEI measures outcomes following a single consultation. For many patients, outcomes will become apparent only after longer episodes of
care. Such outcomes may be multi-layered, capturing aspects of enablement, resilience, symptoms and function, and health perceptions. Without a generic PROM that captures such domains, it is impossible to properly assess the outcome of new primary care service configurations from a patient’s perspective.

The Primary Care Outcomes Questionnaire (PCOQ) was developed in this context, through a rigorous process underpinned by a conceptual model of outcomes which included patient health status and ability to impact health status (see Figure 1). We firstly interviewed patients and clinicians to establish outcomes which both groups sought to achieve within this framework. We then consulted with patients, clinicians and academics in a Delphi study, to identify outcomes most relevant to health and able to be addressed in primary care. We then developed and tested, through cognitive interviews, a PROM which addressed these outcomes. This included health status outcomes; internal features of health empowerment (e.g., understanding and ability to self-care); external features (including having access to support and availability of good healthcare); and outcomes about patient’s health perceptions (e.g., health concerns and confidence that they are dealing with their health conditions).

The current study reports on the final stage of this process, which aimed to establish the psychometric properties of the PCOQ in primary care patients.

Methods

Sample and Procedures

Waiting room recruitment was chosen as an appropriate method for recruiting patients seeking primary healthcare. Adult patients were approached in waiting rooms, prior to consultations, in five practices in south-west England, and those consulting a GP for themselves were invited to participate. These included a mix of urban/rural and affluent/deprived areas with patients from...
different ethnic backgrounds. Participants self-completed the PCOQ and questions about patient characteristics as they waited for a consultation. Because the baseline questionnaire needed to be short enough to be completed in waiting rooms, we collected comparator questionnaires at follow-up only. Patients were asked to take home a second copy of the PCOQ and comparator PROMs for completion one week later and return via post. A follow-up email reminder or text was sent after 5 days.

**Measures**

The PCOQ contains 27 items scored on a 5-point unipolar adjectival scale (no problems to extreme problems). The scale wording varies according to attribute, as determined by the qualitative study.(26) For example, the item “how much are you currently affected by pain or discomfort” is anchored at “not at all” and “extremely”. In contrast, the item “how much do you understand your health problems” is anchored at “I understand as much as I want” and “I understand very much less than I want”. (See Supplemental File 1). Patient characteristics were collected at baseline, and seven comparator measures at follow-up. These were: the EQ-5D-5L(29), the Patient Activation Measure (PAM)(30), a single item on likelihood of recommending GP(31), the last appointment score(31), a single item on support for long-term conditions(31), the Illness Perception Questionnaire (IPQ) (32), and a single item on Change in Main Problem(33). These measures are described in Supplemental File 2.

**Analysis**

Psychometric testing of a multi-item PROM includes evaluations of feasibility, structural validity, internal consistency, construct validity and responsiveness.(13, 34) In this study, feasibility was assessed by the amount and pattern of missing data, readiness of patients to complete the measure, and by response rate between baseline and follow-up. Structural validity was tested using
exploratory factor analysis with principal axis factoring on fully completed questionnaires. (35) This is a statistical technique used to reduce a larger number of items into a smaller number of common factors that reflect shared variance. (35) The number of factors extracted was decided by a combination of Kaiser’s rule (eigenvalues greater than one), the scree plot, and by interpretability of domains. Internal consistency reliability of each domain was assessed using Cronbach’s alpha. (25) Construct validity was explored by testing prespecified hypotheses about the relationships between PCOQ domain scores and comparator PROMs, using Spearman correlations. (25) Finally, responsiveness was tested by comparing Glass’ delta for patients expected to improve versus those expected to remain unchanged, based on patient responses to the item asking about change (thinking about the main problem you consulted your GP with at your recent appointment, is this problem: completely better, much better, better, slightly better, same, slightly worse, worse, n/a). Glass’ delta is a repeated measures effect size, calculated as the mean change in scores (baseline to follow-up) divided by the standard deviation of scores at baseline.

Findings

Feasibility

The PCOQ was accepted by 718 people in the waiting room, and finished by 602 (84%). Missing data at baseline varied between 1% and 7% per item, with 2.5% missing overall. 512 questionnaires (85%) had no missing data. Of the 602 patients completing the PCOQ at baseline, 264 (44%) completed the follow-up questionnaire.

Table 1 shows patient characteristics at baseline and at follow-up. Older patients (apart from the 75+ group) had higher response rates at follow-up.
Structural validity

Kaiser’s rule suggested a four or five-factor solution, and the scree plot a two-factor solution (see Supplemental File 3). Exploratory factor analyses were thus conducted using the complete baseline data (n=512) for solutions ranging from one to five factors. The four-factor solution with oblique (promax) rotation(35) provided the most interpretable simple structure (see Table 2). Three items which did not load highly on any factor were removed. Two of these items were related to health concerns, a construct also reflected in other items. The third was on medication side-effects. The obliquely rotated factors were moderately correlated (0.29 – 0.51). The factors were labelled and described as follows:

Factor 1:  **Health & Well-being**: Measures overall health status, including symptoms, effects of symptoms on life and health concerns.
(8 items)

Factor 2:  **Health Knowledge & Self-Care**: Measures health knowledge, and patients’ ability to self-care and manage symptoms.
(4 items)

Factor 3:  **Confidence in Health Provision**: Measures patients’ confidence in their healthcare providers and ability to access good healthcare.
(6 items)

Factor 4:  **Confidence in Health Plan**: Measures patients’ confidence in their health plan, their adherence to this plan and the level of support they have in managing their health-related problems.
(6 items)

A score was calculated for each domain using a simple average of item scores for each domain (scored 1 – 5). Alternative scores incorporating factor weights were also produced, but converged with the non-weighted scores (r = 0.99) and the simpler method was thus preferred. The PCOQ score distributions are shown in Table 3. There is evidence of a ceiling effect in some domains. For example, all patients scoring in the top quintile for both Health Knowledge & Self-Care and Confidence in Health Provision scored at the ceiling. However, the ceiling effect for Confidence in
Health Provision was lower than the ceiling for the main comparator, Last Appointment Score (25% as opposed to 36%). Similarly, the ceiling effect for Health & Well-being was lower than the EQ-5D (12% at the ceiling as opposed to 20% for the EQ-5D).

**Internal Consistency**

Cronbach’s alpha was above the standard of 0.7 (34) for each domain as follows: Health & Well-being ($\alpha=0.88$), Confidence in Health Provision ($\alpha=0.95$), Health Knowledge & Self-Care ($\alpha=0.85$), Confidence In Health Plan ($\alpha=0.77$).

**Construct validity**

Spearman rho correlations of PCOQ domain scores with comparator PROMs (Table 4) indicate convergent and discriminant construct validity (25): domain scores correlate with PROMs as expected on conceptual grounds (for example Health & Well-Being with EQ-5D, $\rho=0.75$), and less strongly with the other PROMs that reflect different constructs (for example Health & Well-Being with PAM-13, $\rho=0.28$).

**Responsiveness**

We hypothesised that patients responding positively to the Change in Main Problem item (slightly better to much better) would have a positive Glass’ delta ($>0.2$) for Health & Well-being, Health Knowledge & Self-Care and Confidence in Health Plan, if they indicated problems in these domains when they attended their GP (patients were excluded if they were at the ceiling at baseline, as this indicated that they had no problem in that domain when attending). We also expected patients who were perfectly satisfied on the Last Appointment Score to have a positive Glass’ delta for the domain.
Confidence in Health Provision. We only used the top score because the Last Appointment Score had a strong positive skew, which meant that using, for example, the top two categories would have included most patients.

The effect sizes for patients expected to improve are positive, and small to moderate in magnitude with confidence intervals excluding zero (indicating statistical significance at the 0.05 level). Effect sizes for patients expected to stay the same all approach zero (see Table 5). The sample of patients reporting a deterioration on the change item was very small and was therefore not analysed.

Discussion

Summary
Following a rigorous process of development, we have tested the psychometric properties of a PROM designed to capture outcomes that patients want to obtain from primary care and which doctors seek to deliver. This meets a need for an instrument that can determine the effects of alternative forms of primary care, where patients have various problems and reasons for consultation.

Strengths and Limitations
The PCOQ has advantages over existing PROMS, and meets recommended standards for psychometric testing in this sample of primary care patients.(34) Strengths of the study include successful data collection, a simple factor structure with good construct validity, and a prospective design which enabled tests of responsiveness. The study also had limitations.

Some patients did not accept a questionnaire, and 16% were called to an appointment before completing the baseline PCOQ. Although efforts were made to obtain a mix of urban and rural practices, with different deprivation scores, data were collected in a relatively small number of practices. The sample was representative of general practice consultation in terms of gender(36) and
number of long-term conditions (16). However, patients over 75 were slightly underrepresented (37), perhaps because housebound patients were excluded or more declined. Response rates from ethnic groups may have been affected by the lack of interpretation facilities. Only 44% of patients returned the follow-up form, and this differed systematically by age (although not by gender, long-term conditions or ethnicity). This response rate is comparable to similar postal questionnaires (38) and was anticipated by our protocol. Baseline and follow-up scores were compared only for the final responsiveness tests. Furthermore, this responsiveness analysis was based on comparing patients who were separated into two groups based on a change score, and there is no reason to assume different numbers of non-responders across groups. However, we recognise that selection bias may have been introduced. As with comparable questionnaires (39, 40), most items and domain scores were positively skewed. The factor solution had high uniqueness for some items. This can indicate that the item is not strongly related to others (35) but because of the important content of these variables (e.g. pain, adherence), we chose to include them. The Cronbach’s alpha of 0.95 for the Confidence in Health Provision factor may indicate some item redundancy (25) and future validation might further reduce the items. A final limitation relates to the lack of a gold standard for measuring change. The Change in Main Problem and Last Appointment Score were used as proxies, but these are imperfect measures. However, this is a necessary limitation of developing a PROM with a new and unique set of constructs, and the evidence for responsiveness is akin to the concept “construct responsiveness” as described by the COSMIN group (41, 42).

**Comparison with existing literature**

Taken together, the four domains of the PCOQ have commonalities with the concept of health capability, defined as combining health agency (an individual’s ability to achieve health goals and act as agents of their own health) and health functioning (the outcome of actions to maintain or
The four domains are scored separately, and each domain has benefits over other generic PROMs.

The Health & Well-being domain, which includes physical / emotional symptoms, life effects and health concerns, was the most responsive, and had a lower ceiling effect than the EQ-5D, another measure of health status. However, we did not carry out a head-to-head comparison of responsiveness, and this is a future research requirement. Similarly, the responsiveness of the Health Knowledge & Self-Care domain should be compared with similar measures such as the PAM-13. While it showed a stronger ceiling effect than the PAM-13, this domain includes areas often receptive to intervention, such as patient understanding of health problems, while excluding areas, captured by PAM-13, which may be less responsive, such as figuring out solutions to new health problems. The Confidence in Health Plan domain refers to patients trusting and following their health plan, and having support to enable this. This is a broad construct that subsumes different concepts, and no existing PROM is readily comparable. Lastly, the Confidence in Health Provision domain includes concepts similar to those normally found in patient-reported experience measures (PREMs) as opposed to PROMs, such as whether the clinician listens. However, unlike a PREM, which refers to perceptions of a particular consultation, the PCOQ refers to patients’ current levels of confidence in these aspects of their health providers. These are best viewed as outcomes rather than experiences. That these levels are amenable to change following a consultation is a valuable result of this study.

**Implications for research and/or practice**

This study has demonstrated that the PCOQ is valid, internally consistent and responsive among this sample of primary care patients. Because this study represents the first validation of the PCOQ, we recommend usage alongside other PROMs until properties are confirmed. The PCOQ was specifically developed to test the benefits of service-level interventions in primary care, and thus fills an
important gap in the literature. The alternative to the PCOQ is multiple instruments measuring four different constructs. This would not only require a longer questionnaire, but also usage of instruments not designed to measure outcomes that primary care patients seek. No existing PROM covers the PCOQ unique construct. It therefore offers a timely opportunity to enhance research and policy making in primary care during a period of high demand for new interventions in this area.

**Additional information**

**Funding Information**

This study was funded by the National Institute of Health Research School for Primary Care Research (NIHR SPCR). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. Mairead Murphy was partly funded by the Avon Primary Care Research Collaborative.

**Ethical Approval**

Ethical approval was granted through the National Health Service Research Ethics Committee process. Approval was granted by London Central Committee, reference 14/LO/2036.

**Competing Interests**

The authors have no potential conflict of interest or competing interests. The study design, data collection, analysis and interpretation of data was undertaken entirely by the authors, and the study sponsors had no role in this.
Acknowledgements

The authors would like to thank all the participants in this study, in particular the patients who completed the questionnaires and the primary care centres that allowed us to administer it. We would also like to thank the Bristol Primary Care Research Network for assisting with recruiting the participants, and the National Institute of Health Research School for Primary Care (NIHR SPCR) for funding the research. The NIHR SPCR is a partnership between the Universities of Bristol, Cambridge, Keele, Manchester, Newcastle, Nottingham, Oxford, Southampton and University College London. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

The PCOQ, and associated scoring files and instructions are available from the University of Bristol Centre for Academic Care website: http://www.bristol.ac.uk/primaryhealthcare/resources/

Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>EFA</td>
<td>Exploratory Factor Analysis</td>
</tr>
<tr>
<td>EQ-5D</td>
<td>European Quality of Life-5 Dimensions</td>
</tr>
<tr>
<td>ISOQOL</td>
<td>International Society for Quality of Life Research</td>
</tr>
<tr>
<td>MYMOP</td>
<td>Measure Yourself Medical Outcomes Profile</td>
</tr>
<tr>
<td>NHS</td>
<td>National Health Service</td>
</tr>
<tr>
<td>NIHR</td>
<td>National Institute for Health Research</td>
</tr>
<tr>
<td>PAM-13</td>
<td>Patient Activation Measure (13 item version)</td>
</tr>
<tr>
<td>PCOQ</td>
<td>Primary Care Outcomes Questionnaire</td>
</tr>
<tr>
<td>PEI</td>
<td>Patient Enablement Instrument</td>
</tr>
<tr>
<td>PREM</td>
<td>Patient-reported experience measure</td>
</tr>
<tr>
<td>PROM</td>
<td>Patient-reported outcome measure</td>
</tr>
<tr>
<td>SF-36</td>
<td>Short Form 36</td>
</tr>
</tbody>
</table>
References