Patient preferences for patient participation: Psychometric evaluation of The 4Ps tool in patients with chronic heart or lung disorders

Kristina Luhr, Ann Catrine Eldh, Ulrica Nilsson and Marie Holmefur

The self-archived postprint version of this journal article is available at Linköping University Institutional Repository (DiVA):

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-146102>

N.B.: When citing this work, cite the original publication.

Luhr, K., Eldh, A. C., Nilsson, U., Holmefur, M., (2018), Patient preferences for patient participation: Psychometric evaluation of The 4Ps tool in patients with chronic heart or lung disorders, *Nordic journal of nursing research*, 38(2), 68-76. <https://doi.org/10.1177/2057158517713156>

Original publication available at:

<https://doi.org/10.1177/2057158517713156>

Copyright: SAGE Publications (UK and US)

<http://www.uk.sagepub.com/home.nav>

Title: Patient Preferences for Patient Participation – psychometric evaluation of The 4Ps tool in patients with chronic heart or lung disorders

Background

Patient participation is declared by The World Health Organization to be a key issue for good care.[1] In Sweden, healthcare professionals are required to provide conditions for patient participation,[2, 3] and recent legislation further emphasizes the patient's right to participation, involvement and influence over their own care.[3] However, different perspectives of the concept of “patient participation” are applied by different partakers,[4] and when the patients’ preferences for participation are not recognised, the healthcare staff’s assumptions[5] and/or the culture of the healthcare facility[6, 7] have been found to drive the planning and provision of care. Thus, it is important to present opportunities for the individual patient to voice his or her preferences for participation to fulfil the intentions of the legislation.

There are some instruments available that retrospectively evaluate patients’ experiences of participation, involvement and influence over their own care.[8-13] However, a prospective approach can be beneficial for the individual patient, as well as be of help to the healthcare staff in the implementation of patient participation in practice; a tool that allows the patient to share his or her preferences for participation, combined with the opportunity to evaluate patient participation, can aid in the dialogue and provide a basis for individualised care plans. The Patients Preferences for Patient Participation tool (The 4Ps) [14] was developed in 2009 and 2010 based on the above requirements, allowing the individual patient to depict, prioritize, and later evaluate his or her participation in three separate sections. Thus, The 4Ps

is a clinical tool, intended for patients with planned or expected recurrent contacts with care, to phrase and share preferences for their participation with the health professionals, thereby offering the professionals an opportunity to understand and provide for patient participation from the patient's perspective. Further, The 4Ps offers the patient an opportunity to evaluate his/her experience of patient participation. In a dialogue between health professionals and patient, the latter's experience can be considered; if and to what extent the experiences match the preferences for patient participation can be evaluated. A potential mismatch can be jointly examined, for a mutual dialogue on what lesson is learned, for both the patient and the health professional/s. Given that The 4Ps tool works for comparisons between prioritizations and evaluations at individual patient and/or group level, the tool has potential to aid both clinic and science as regards patient participation and person-centred care issues.[15]

Previously, The 4Ps tool was tested as regards content validity, response process, and acceptability in a qualitative study with content and construct experts: scientists engaged in patient participation, and patients with chronic heart failure (CHF) or chronic obstructive pulmonary disease (COPD) took part in the research. [14] The findings showed all items to be relevant and to correspond to the concept of patient participation. Further, the tool was considered to have an appealing scope, structure and format. Yet, minor amendments were recommended and additional evaluation was necessary;[16] thus, the purpose of this study was to further evaluate aspects of reliability and validity of The 4Ps in patients subject to chronic heart or lung disease.

Methods

Design

The psychometric properties of The 4Ps tool were investigated by evaluating internal scale validity and test-retest reliability. For study purposes, the potential for using The 4Ps for a dialogue between the patient and health professional about preferences for and experience of participation was excluded, and data was collected for the psychometric evaluation only.

Ethical approval was given by the Regional Ethical Review Board in Uppsala, Sweden (number 2011 / 032).

The 4Ps tool

The 4Ps tool consisted of three sections, for the patient to 1) depict, 2) prioritize, and 3) evaluate patient participation, with 12 pre-set items reiterated in each section.[14] The items were formed in four aspects of patient participation: *having dialogue with health-care staff*, *sharing knowledge*, *partaking in planning* and *managing self-care*, with three items in each aspect (table 1). In section 1 of the 4Ps, the patient was guided to conceptualize patient participation by ticking the items that convey his or her idea of the concept, in order to encourage the patient to reflect what patient participation means to him/her. In section 2, the patient considered to what extent each item is a priority for participation in his/her upcoming healthcare interaction. This section used a scale with the four alternatives of ‘*completely unimportant*’, ‘*less important*’, ‘*quite important*’, and ‘*very important*’. Sections 1 and 2 were to be completed by the patient at the same occasion. Section 3 guided the patient to evaluate to what extent he or she has experienced participation, using a 4-point scale including ‘*not at all*’, ‘*to a limited extent*’, ‘*to some extent*’, and ‘*completely*’.

//Table 1 about here//

Sample

The study used a strategic sample; individuals with CHF or COPD were recruited through primary healthcare centres and outpatient CHF clinics in both urban and rural areas in a typical Swedish region. Inclusion criteria were: the ability to respond to a questionnaire in Swedish, and being planned for a minimum of two healthcare contacts within the next 12 months. For the study, 50 patients diagnosed with COPD and 60 patients diagnosed with CHF were recruited. Data were collected from October 2011 to August 2013.

Procedure

Registered nurses (RNs) at each unit were engaged to identify and include patients. For each patient fulfilling the inclusion criteria, the RN sent written information about the study prior to a planned outpatient visit or provided the written information during a visit. During the scheduled visit, each patient was given a brief oral recap of the information along with a question about their willingness to participate. Patients agreeing to take part gave their written consent and then completed sections 1 and 2 of The 4Ps tool individually, placed the completed tool in a prepaid envelope, and mailed it to the research team. Two weeks after the first completion of sections 1 and 2, a retest of sections 1 and 2 was sent to the patient. Section 3 was sent to the patient once the date of the second of the two additional planned healthcare contacts had passed. In addition, a retest for section 3 was sent to the patient within two weeks after the first completion of section 3. For each section of the tool and the retests, reminders were sent after two and four weeks, respectively.

Statistical analysis

Responses in section 1 were managed as dichotomous data in the statistical analysis, that is, ticked items were treated as 'yes' and non-ticked items as 'no'.

Internal scale validity To evaluate internal scale validity, and to determine two aspects of reliability, the Rasch measurement model was used to analyse the test data of each section in The 4Ps tool. From the family of Rasch models, the dichotomous model was chosen in section 1. In section 2 and 3 with 4-category scale, the polytomous Rating Scale Model was chosen because all items were designed to share the same rating scale.[17] In a Rasch analysis, persons and items are hierarchically ordered along a common measurement line, and scores are transformed by a logarithmic transformation to linear measures and the common unit of measurement for persons and items is “log-odds-probability-units” (logits).[17]

The functioning of the 4-category rating scales of sections 2 and 3 was scrutinized (in section 1, items are dichotomous and thus not considered as a rating scale). The criteria for optimal rating scale functioning was a minimum of 10 responses in each scale step, an outfit mean square (MnSq) of <2 for each category and threshold, and monotonic advances in scale step and threshold measures.[18] Item goodness-of-fit to the Rasch model was investigated, and acceptable item fit to the construct was set to an infit MnSq of residuals between 0.6 and 1.4 (the expected value was 1) in combination with a standardized z-statistic within the range of -2.0 to $+2.0$. [19] The weighted infit item statistics were chosen because they are less affected by extreme responses by persons whose measure is far from the item measure than the unweighted outfit statistics.[17] Person fit was also investigated using the same criteria as previously.[20, 21]

Item and person reliability coefficients were calculated.[17] The item reliability coefficient reflects to what extent the hierarchy of the items is applicable to another set of respondents. The person reliability coefficient indicates the replicability of person ordering, if this group of

persons was given a parallel set of items and is analogous to the traditional coefficient alpha.[17] Benchmarks were set to 0.91-0.94 as very good, 0.81-0.90 as good, 0.67-0.80 as fair and <0.67 as poor.[22] Targeting, i.e. to what extent item and person measure levels match, was investigated by comparing mean logit measures for items and persons, with the set benchmark of <0.5 logits, and by visual inspection of the person-item map of sections 2 and 3 with 4-category scales.[17]

Test-retest reliability For test-retest reliability, intraclass correlation coefficients (ICC) were calculated on person measure.[23] Persons who answered less than six items in either test or retest were excluded from the analysis. The person measure for each person and section was obtained from the Rasch analysis in the unit of logits and then transformed to the more easily interpreted “4Ps units” ranging from 0-100 for each subscale.[17] The ICC (agreement) was calculated, based on one-way analysis of variance (ANOVA), as the reliability coefficient for each section.[23] An ICC of 0.70 was considered to be sufficient.[24] To evaluate test-retest reliability on an item level, the kappa coefficient, which assesses agreement beyond chance, [25] was calculated for each item in each section. For sections 2 and 3, with 4-category scales, the quadratic-weighted kappa coefficient [26] was calculated. To adjust for skewed proportions in the responses, which give a low kappa despite high agreement, a prevalence- and bias-adjusted kappa (PABAK) [27] was calculated. As benchmark, for all kappa calculations [28], a coefficient of 0.81-1.00 was considered almost perfect, 0.61-0.80 substantial agreement, 0.41–0.60 moderate agreement, 0.21–0.40 fair agreement, and 0.00–0.20 slight agreement. [29] For descriptive purposes, the per cent agreement, which includes agreement due to chance, was calculated between test and retest.

A 95% confidence interval (CI) was used when applicable. Rasch analyses were conducted with the Winsteps® Rasch measurement computer program (Version 3.75.1) (Copyright 2009 John M. Linacre). The SPSS version 21 software was used to generate descriptive statistics and to perform ICC calculations. Kappa and weighted kappa were calculated with the online software VassarStats.[30] The 95% CIs for weighted kappa in sections 2 and 3 were calculated manually in the software program STATA release 11. PABAK was calculated using the online Pabak-OS calculator.[31]

Findings

Of the 110 patients who agreed to take part in the study, 98% responded to section 1 and 2, and 88% responded to section 3. For the retests, 93% and 89% responded to sections 1 and section 2, respectively. Due to an administrative error, the retest of section 3 was only sent out to 39 participants of whom 31 (80%) responded (Table 2).

// Table 2 about here //

Internal scale validity

The rating scale analysis, performed in sections 2 and 3, showed that the threshold and category measures increased for each scale step, the outfit MnSq was <2 and the observed counts were >10 for all response alternatives. Thus, all criteria for rating scale functioning were fulfilled. The entire rating scale was employed in section 3 more so than in section 2 with 82% of the ratings in the two most positive alternatives in section 3, compared to 94% in section 2.

All items in section 1 and section 2 showed acceptable goodness-of-fit (Table 3). In section 3, where the patient evaluates participation, 10 out of 12 items showed item fit. Item 9 (infit MnSq 1.59, Zstd 3.40) and item 10 (infit MnSq 1.80, Zstd 3.80) showed misfit, i.e. those items had many unexpected responses in relation to the Rasch assumptions. The person goodness-of-fit analysis showed 99% fit in section 1, 96% fit in section 2, and 95% fit in section 3, *i.e.*, 5% or less showed misfit in each section. Thus, the group of patients responded in a way that fit the Rasch model assumptions and was sufficiently homogenous (table 3).

// Table 3 about here //

The item reliability of section 1 was 0.93 (very good), and the person reliability was 0.65 (poor). In section 2 the item reliability was 0.94, which is very good, and the person reliability was 0.69, which is fair. For section 3, the item and the person reliability were 0.93 (very good) and 0.80 (fair), respectively.

The visual inspection of the person-item map of sections 2 and 3 showed a ceiling effect in both sections, more pronounced in section 2. The difference between the mean item measure and the mean person measure was 3.64 logits in section 2 and 1.92 in section 3, thus considerably higher than the criterion of <0.5 logits (Figure 1). The actual ceiling effects were 36% (n=39), 19% (n=20), and 5% (n=5), in the tests of the sections respectively.

// Figure 1 about here //

Test-retest reliability

The reliability coefficient (ICC), for section 1 was 0.56 (CI=0.39-0.70), and for section 2, 0.56 (CI=0.40-0.70), which is not considered to be sufficient. Section 3 was found stable, with an ICC of 0.82 (CI=0.65-0.91).

On item level in section 1, the PABAK reliability coefficients ranged between 0.45 and 0.76, with half of the items reaching substantial agreement, and the other half moderate agreement (Table 4). Kappa coefficients in this section were 0.12-0.52 and as indicated by the substantial difference between PABAK and Kappa coefficients the responses were unevenly distributed. In section 2, the PABAK coefficients were 0.48–0.77, with seven items reaching substantial agreement (PABAK 0.64-0.77, K_w 0.27-0.62), and the remaining five items moderate agreement (PABAK 0.48-0.59, K_w 0.30-0.56) (Table 4). In section 3, the PABAK coefficients ranged from 0.46 to 0.91 (K_w 0.31-0.89) with item 3 showing almost perfect agreement, item 6 substantial agreement; and seven items moderate agreement (PABAK 0.46-0.60, K_w 0.31-0.75). The remaining three items reached fair agreement (PABAK 0.26-0.36, K_w 0.51-0.64), all derived from the aspect *partaking in planning* (Table 4).

// Table 4 about here //

Discussion

The psychometric evaluation of The 4Ps performed in this study showed inspiring results, and points out some areas where the tool can be improved, along with imperative implications for use in clinical practice and in research.

The Rasch analysis illuminated how The 4Ps is used by the participants. The most obvious result was the skewed responses in the rating scale and targeting analyses. The fact that many of the items were highly prioritized and evaluated is a positive aspect of the instrument in the sense that it indicates that all items are relevant for patient participation.[28] The downside of this is a weaker sensitivity to change because a true change in the patients preferences or evaluations may not be detected.[24] This ceiling effect was most prominent in the dichotomised section 1, where the patients were instructed to tick items that they defined as participation. Since all items included patient participation in general terms,[14] a ceiling effect was expected and not a problem. In section 2, with a ceiling effect of 19 %, providing a response alternative with a value higher than '*very important*' might further support distinguishing which item or items that are a priority to the patient in terms of participation. The responses to section 3 were more diverse, and the slight ceiling effect a must since the item '*completely*' cannot be granted a higher value. The item fit evaluation showed that all items in section 1 and section 2, and all but two in section 3, fit the Rasch model; thus, the 4Ps transpires a unidimensional construct.[17] The misfit in item 9 and item 10 in section 3 can probably be attributed to unclear phrasing in item 9, and the item 10's specific examples of self-care not corresponding to living with CHF or COPD. This would be consistent with the findings in the earlier qualitative validation study.[14] We suggest rephrasing item 9 to emphasize that this aspect of patient participation connotes an interaction with healthcare professionals, in line with Swedish healthcare legislation[3], and removing the examples of self-care presented in item 10. Further, the reliability results from the Rasch analysis showed

very good item reliability, and poorer person reliability. The poorer person reliability is most likely due to the ceiling effects, an effect discussed above along with potential solutions.

The test-retest reliability in sections 1 and 2, illustrated by ICC coefficients, did not reach the set criterion. One reason for the poor ICC may be that only the upper half the 0-100-scale was used for sections 1 and 2, and because the ICC is a function of the within- and between-person variance, a small between-person variance tends to deflate the ICC. Nevertheless, the poor agreement between test and retest in section 1 and 2 is also noticed in particular items. On the item level, PABAK and Kappa coefficients ranged between moderate to substantial agreement. When interpreting the test-retest results, the purposes of the different sections of The 4Ps tool should be considered. Section 1 is designed to merely guide the patient into the subject of participation and in section 2, the patients are asked to prioritize the importance of the items to patient participation. Although frequently used in policy documents in healthcare, [32, 33] 'patient participation' is an abstract concept and it is possible that lay people have not reflected upon its meaning. Consequently, the first test session might have led to further reflections on the issue of participation and the role of the patient, and this might have affected the retest of sections 1 and 2.[28] Furthermore, the patients' preferences might vary and individuals might change their mind regarding which aspects of the concept are the most prominent at a particular point in time.[34] These results show that patient preferences for participation may vary over time; thus, in clinical practice patients' preferences for participation need to be addressed recurrently. This can, together with the prominent ceiling effect in section 2, explain the modest test-retest reliability in section 1 and 2, in both item level and in person measure.

The purpose of section 3, on the other hand, is to evaluate patient participation, and for this purpose high test-retest reliability is essential. In section 3, the test-retest reliability was found to be more stable when using person measures, whereas it varied between items. One potential reason for the less-favourable results may be that the patients could have had additional healthcare contacts during the test-retest period of 14 days. In such case the patients' evaluation of patient participation would not be of the same healthcare interaction, potentially affecting the re-test. Further, all-in-all what was evaluated was the recall of the individual's experience of participation in healthcare contacts. Apparently, the recall of the verbal contact with healthcare staff was more stable (items 3 and 6), while the memory of other aspects was less stable. The lowest test-retest reliability was found in the aspect *Partaking in planning*. Possibly, the low reliability was due to that the patients had no, or limited, experience of planning to relate to, as reported previously in the qualitative validation study.[14] Despite low stability for some items, we find no reason to reduce or modify items in the aspect *Partaking in planning* since planning of care is a well-defined aspect of patient participation, recognised in both research and policies.[14, 35] Instead, these results need to be taken into consideration when using The 4Ps for evaluation purposes. However, when calculating test-retest reliability from person measures, section 3 was found more stable. This illuminates the advantage of using person measures for research purposes which in this case seem to even out the differences between items. Therefore, the person measures are preferable to individual items in research cases. Further, when performing a test-retest evaluation, there has to be a sufficiently short interval between the test and the retest so that the measured phenomena is not likely to have changed.[28] In this study, the time frame of two weeks with an additional option to respond to reminders after four or six weeks might have been too long. Most other instruments measuring patient participation have not been evaluated for test-retest reliability.[8, 11-13] The exception is one patient-participation instrument that demonstrated

acceptable test-retest reliability with ICC ranging from 0.59 to 0.93 on an item level.[10]

However, that instrument evaluated patient's participation in emergency departments for one specific occasion only, whereas section 3 in The 4Ps evaluates participation in on-going care for patients suffering from a chronic condition. Considering the differences in healthcare processes and settings, in addition to different reasons for interacting with healthcare services, we suggest that a tool reflecting on-going care like The 4Ps will not be as stable as a tool evaluating one specific healthcare event.

The clinical implications of this study are that the patients' evaluation in section 3 is suggested useful for recurrent dialogues between the patient and healthcare professional. Further, with proviso that the reliability of section 2 is improved, the patients' prioritization (section 2) of participation can be compared to the patients' evaluation (section 3), again in a dialogue between the patient and the healthcare professional. The patients' prioritization in section 2 could also be compared with previous prioritizations. This process would aid healthcare professionals to better understand patient participation considering both patients' preferences and patients' experiences. Thus, it can provide a means for understanding and implementing patient participation policies.[3, 33] In clinical practice, the comparisons can supposedly be performed item by item at an individual level, or group level. Whereas in research, we propose The 4Ps to be applied for evaluation purposes: with suggested amendments in section 2, a correspondence between sections 2 and 3 can be determined, *i.e.* taking both patients' preferences and experience of patient participation into account. Further, section 2 can be used to evaluate whether patients change their prioritizations over time or after an intervention.

Legislations and policies have a special significance in healthcare, a sector based on knowledge and evidence.[3] However, policies can be based on standards and culture, and can be multifaceted and abstract. Yet, health professionals and organizations are to interpret and implement policies into practice. Regardless of the increasing awareness of individuals' right to autonomy in diverse interactions with society, there is still a need to implement means that facilitate patients' right to participate in healthcare. We advocate that innovations like The 4Ps, with opportunities for the patient and staff to interact, may serve patient participation from a patient perspective.[36] However, further studies are needed with regards to a favourable implementation of The 4Ps in clinical practice.

Methodological considerations

The 4Ps tool was developed for clinical use to provide a means for patients to share their preferences for participation with healthcare professionals in clinic. In addition, the responses can be used in research to increase knowledge on patient participation in subgroups of patients.[16] The psychometric evaluation was performed in two levels: an item level evaluation, where the application area primarily is found in clinic, and an evaluation performed on person measures, exclusively for evaluation for scientific purposes. A suitable statistic for evaluating agreement in ordinal level data, in this study in the item level, is the kappa analysis. However, kappa is dependent on equal distributions of answers across all response options in order to give fair coefficients. In the case of The 4Ps, a higher frequency of answers at the higher end of the rating scale is to be expected. This skewness in responses inevitably gives a lower kappa coefficient despite fair agreement, called "the kappa-paradox" and leads to difficulties in interpreting the kappa coefficients and drawing conclusions on reliability.[37] Therefore, to provide a more comprehensive understanding of the agreement we calculated PABAK which compensates for "the kappa-paradox".[27, 38] The sample in

the retest of section 3 was considerably smaller than planned. This affected the CI of the reliability coefficients, thus the CI became wide.[28] The low response rate was not mainly due to patients not wanting to answer the retest, but rather, due to error, only a few patients were given the opportunity to respond.

Conclusion

The 4Ps tool provides a means to increase the understanding of patient participation in clinical practice, providing a structure for the individual patient to become more involved in his or her care. As such, it is a potential aid in implementing policies on patient participation.

Furthermore, The 4Ps can be applied to increase general knowledge of patient participation. When psychometrically evaluated, The 4Ps tool was found to have reasonable validity, and varied reliability. Based on the findings, we suggest the rating scale in section 2 to be modified to improve sensitivity and reliability, as well as rephrasing of two items to improve goodness-of-fit. Following amendments, The 4Ps tool can be suggested for recurrent use in dialogues in everyday healthcare, as a basis for care planning together with patients subject to chronic heart or lung disease in primary and/or outpatient healthcare, and for scientific purposes in similar target groups.

References

1. World Health Organization. *The Health for All policy framework for the WHO European Region: 2005 update*. Copenhagen: WHO, 2005.
2. The Swedish Parliament. *Hälso- och sjukvårdslag (Health Care Act) (1982:763)*. Stockholm: The Swedish Parliament, 1982.
3. The Swedish Parliament. *Patientlag (Patient law) (2014:821)* Stockholm: The Swedish Parliament, 2014.
4. Eldh AC, Ehnfors M and Ekman I. The meaning of patient participation for patients and nurses at a nurse-led clinic for chronic heart failure. *Eur J Cardiovasc Nurs*. 2006; 5: 45-53.
5. Florin J, Ehrenberg A and Ehnfors M. Patient participation in clinical decision-making in nursing: A comparative study of nurses' and patients' perceptions. *J Clin Nurs*. 2006; 15: 1498-508.
6. Aasen EM, Kvangarsnes M and Heggen K. Perceptions of patient participation amongst elderly patients with end-stage renal disease in a dialysis unit. *Scand J Caring Sci*. 2012; 26: 61-9.
7. Frank C, Asp M and Dahlberg K. Patient participation in emergency care - a phenomenographic study based on patients' lived experience. *Int Emerg Nurs*. 2009; 17: 15-22.
8. Lindberg J, Kreuter M, Person LO and Taft C. Patient Participation in Rehabilitation Questionnaire (PPRQ)-development and psychometric evaluation. *Spinal Cord*. 2013; 51: 838-42.
9. Lund ML, Nordlund A, Bernspang B and Lexell J. Perceived participation and problems in participation are determinants of life satisfaction in people with spinal cord injury. *Disabil Rehabil*. 2007; 29: 1417-22.

10. Frank C, Asp M, Fridlund B and Baigi A. Questionnaire for patient participation in emergency departments: development and psychometric testing. *J Adv Nurs*. 2011; 67: 643-51.
11. Arnetz JE and Arnetz BB. The development and application of a patient satisfaction measurement system for hospital-wide quality improvement. *Int J Qual Health Care*. 1996; 8: 555-66.
12. Arnetz JE, Hoglund AT, Arnetz BB and Winblad U. Development and evaluation of a questionnaire for measuring patient views of involvement in myocardial infarction care. *Eur J Cardiovasc Nurs*. 2008; 7: 229-38.
13. Wilde Larsson B and Larsson G. Development of a short form of the Quality from the Patient's Perspective (QPP) questionnaire. *J Clin Nurs*. 2002; 11: 681-7.
14. Eldh AC, Luhr K and Ehnfors M. The development and initial validation of a clinical tool for patients' preferences on patient participation--The 4Ps. *Health Expect*. 2015; 18: 2522-35.
15. Price B. Exploring person-centred care. *Nurs Stand*. 2006; 20: 49-56; quiz 8.
16. American Psychological Association. *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association, 1999.
17. Bond T and Fox C. *Applying the Rasch model : fundamental measurement in the human sciences*. 2 ed. New Jersey: Lawrence Erlbaum Associates, 2007.
18. Linacre JM. Optimizing rating scale category effectiveness. *J Appl Meas*. 2002; 3: 85-106.
19. Wright B and Linacre J. Reasonable mean-square fit values. *Rasch Measurement Transactions*. 1994; 8: 370.
20. Smith RM, Schumacker RE and Bush MJ. Using item mean squares to evaluate fit to the Rasch model. *J Outcome Meas*. 1998; 2: 66-78.

21. Wang WC CT. Item Parameter Recovery, Standard Error Estimates, and Fit Statistics of the Winsteps Program for the Family of Rasch Models. *Educ Psychol Meas.* 2005; 65: 376-404.
22. Fisher W, Jr. Rating scale instrument quality criteria. *Rasch Measurement Transactions.* 2007; 21.
23. Shrout PE and Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull.* 1979; 86: 420-8.
24. De Vet H.C.W, Terwee C.B, Mokkink L.B and D.L. K. *Measurement in Medicine.* Cambridge: Cambridge University Press, 2011.
25. Altman DG. *Practical Statistics for Medical Research.* London: Chapman and Hall, 1991.
26. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull.* 1968; 70: 213-20.
27. Sim J and Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther.* 2005; 85: 257-68.
28. Streiner DL and Norman GR. *Health measurement scales - a practical guide to their development and use.* 4 ed. Oxford: Oxford University Press, 2008.
29. Landis JR and Koch GG. The measurement of observer agreement for categorical data. *Int J Biom.* 1977; 33: 159-74.
30. Vassarstats. VassarStats: Website for Statistical Computation.
<http://www.vassarstats.net/kappa.html> (2001)e , last accessed 1 December 2016)
31. Single Case Research. Single Case Research.
<http://www.singlecaseresearch.org/calculators/pabak-os> (2011, last accessed 1 December 2016)

32. The National Board. *Patientens rätt till information, delaktighet och medinflytande (The patient's right to information, participation and empowerment)*. Stockholm: The National Board, 2003.
33. The National Board. *Din skyldighet att informera och göra patienten delaktig - Handbok för vårdgivare, verksamhetschefer och personal (Your Obligation to Inform and Make the Patient Participate - Handbook for Health Organisations, Managers, and Professionals)*. 3 ed. Stockholm: The National Board, 2012.
34. Hoglund AT, Winblad U, Arnetz B and Arnetz JE. Patient participation during hospitalization for myocardial infarction: perceptions among patients and personnel. *Scand J Caring Sci*. 2010; 24: 482-9.
35. The Swedish Parliament. *Patientdatalag (Patient data law) (2008:355)*. Stockholm: The Swedish Parliament, 2008.
36. Chaboyer W, McMurray A, Marshall A, et al. Patient engagement in clinical communication: an exploratory study. *Scand J Caring Sci*. 2016.
37. Feinstein AR and Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. *J Clin Epidemiol*. 1990; 43: 543-9.
38. Hoehler FK. Bias and prevalence effects on kappa viewed in terms of sensitivity and specificity. *J Clin Epidemiol*. 2000; 53: 499-503.

Figure 1 Variable map. Targeting of sections 2 (1a) and 3 (1b) of The 4Ps tool item measures (to the right in the figures) to The 4Ps tool level of the sample (to the left in the figures). M = mean, S = 1 standard deviation and T = 2 standard deviations.

Table 1. The Patients Preferences for Patient Participation tool (The 4Ps), aspects and items.

Aspect	Item
Having dialogue with health-care staff	1. There are conditions for mutual communication
	2. My knowledge and preferences are respected
	3. Healthcare staff listen to me
Sharing knowledge	4. I get explanations for my symptoms/issues
	5. I can tell about my symptoms/issues
	6. Healthcare staff explain the procedures to be performed/that are performed
Partaking in planning	7. Knowing what is planned for me
	8. Taking part in the planning of care and treatment
	9. Phrasing personal goals
Managing self-care	10. Performing some care myself, such as managing my medication or changing a dressing
	11. Managing self-care, such as adjusting my diet or performing preventive health care
	12. Knowing how to manage my symptoms

Table 2 Demographic information on patients completing the test and retest of the different sections of The 4Ps tool

	Section 1		Section 2		Section 3	
	Test	Retest	Test	Retest	Test	Retest
	(n=108)	(n=100)	(n=108)	(n=97)	(n=95)	(n=31)
Men (n)	60	56	60	55	54	13
Age (years)						
Mean	69	69	69	69	69	69
Min-max	36-89	36-88	36-89	36-88	36-89	55-80
Years the patients had known of their diagnosis						
Mean	4	4	4	4	4	8
Min-max	0 ^a -15	0 ^a -15	0 ^a -15	0 ^a -15	0 ^a -15	0 ^a -15

^a = Recently diagnosed.

Table 3 Item measure and goodness-of-fit statistics of The 4Ps tool.

Section 1, n = 108						Section 2, n = 108					Section 3, n = 95				
Item	Item	Model	Infit	Infit	Total	Item	Model	Infit	Infit	Total	Item	Model	Infit	Infit	
number	measure	S.E	MnSq	Zstd	counts	measure	S.E	MnSq	Zstd	counts	measure	S.E.	MnSq	Zstd	
	logits					logits					logits				
1	−0.90	0.35	1.40	1.90	105	−0.56	0.25	1.06	0.40	91	−0.11	0.18	0.73	−1.80	
2	0.67	0.31	1.07	0.50	98	0.92	0.21	1.01	0.10	89	0.31	0.18	0.61	−2.80	
3	−1.45	0.39	1.19	0.90	101	−1.05	0.27	0.92	−0.40	93	−0.46	0.19	0.71	−1.90	
4	−1.61	0.41	0.68	−1.40	104	−1.06	0.27	0.99	0	93	−0.57	0.20	0.75	−1.60	
5	0.10	0.32	0.89	−0.70	102	−0.08	0.24	0.94	−0.3	93	−1.04	0.21	0.76	−1.50	
6	−1.78	0.42	0.96	−0.10	104	−0.83	0.26	1.16	0.90	91	−0.61	0.20	0.63	−2.50	
7	−0.32	0.33	1.18	1.10	103	−0.24	0.24	1.28	1.50	89	0.95	0.17	1.10	0.70	
8	1.62	0.31	0.92	−0.50	104	1.20	0.20	0.61	−2.7	91	1.27	0.16	1.11	0.80	
9	2.75	0.34	0.93	−0.40	94	1.70	0.19	0.89	−0.6	86	1.20	0.17	1.59	3.40	

10	1.14	0.31	1.01	0.10	103	1.02	0.20	1.27	1.5	89	-0.68	0.20	1.80	3.80
11	0.67	0.31	0.78	-1.50	101	0.45	0.22	1.00	0.1	93	-0.07	0.18	1.15	1.0
12	-0.90	0.35	1.00	0.00	105	-1.47	0.29	0.82	-0.9	91	-0.18	0.19	0.86	-0.90

S.E. = Standard Error.

Infit MnSq = Infit mean square. Weighted mean of the information-weighted standardized residuals.

Infit Zstd = Infit z-score standardized. The infit MnSq value standardized to a t distribution, estimating the statistical significance of the infit misfit.

Table 4 Test-retest reliability results of The 4Ps tool

Item number	Test retest section 1, n = 100			Test retest section 2, n = 97			Test retest section 3, n = 31		
	PABAK	Kappa	Per cent	PABAK	K _w	Per cent	PABAK	K _w	Per cent
	(95% CI)	(95% CI)	agreement	(95 %CI)	(95% CI)	agreement	(95% CI)	(95% CI)	agreement
1	0.65 (0.50–0.80)	0.32 (0.02–0.61)	83	0.71(0.62–0.80)	0.36 (0.16–0.57)	78	0.52 (0.36–0.68)	0.47 (0–1)	64
2	0.48 (0.39–0.65)	0.40 (0.21–0.60)	74	0.66(0.57–0.75)	0.62 (0.17–1)	74	0.51 (0.34–0.67)	0.38 (0.23–0.53)	63
3	0.68 (0.54–0.82)	0.12 (0–0.52)	84	0.69(0.60–0.78)	0.34 (0.15–0.57)	77	0.91 (0.75–1.06)	0.89 (0.25–1)	93
4	0.76 (0.63–0.89)	0.43 (0.13–0.73)	88	0.68(0.59–0.77)	0.36 (0.16–0.56)	76	0.51 (0.36–0.67)	0.39 (0.13–0.65)	63
5	0.60 (0.44–0.76)	0.44 (0.22–0.66)	80	0.59(0.50–0.67)	0.39 (0.27–0.50)	69	0.56 (0.40–0.71)	0.40 (0–0.80)	67
6	0.73 (0.59–0.86)	0.17 (0–0.59)	87	0.64(0.55–0.73)	0.29 (0.09–0.49)	73	0.71 (0.55–0.87)	0.73 (0.07–1)	79
7	0.65 (0.50–0.80)	0.42 (0.16–0.67)	83	0.64(0.55–0.73)	0.27 (0.07–0.47)	73	0.31 (0.15–0.47)	0.60 (0.30–0.90)	45
8	0.45 (0.28–0.62)	0.42 (0.24–0.61)	73	0.51(0.42–0.60)	0.43 (0.14–0.72)	63	0.26 (0.11–0.42)	0.64 (0.40–0.87)	45
9	0.52 (0.35–0.69)	0.52 (0.35–0.69)	76	0.48(0.39–0.57)	0.56 (0.25–0.86)	61	0.36 (0.19–0.52)	0.51 (0.09–0.93)	52
10	0.56 (0.40–0.72)	0.47 (0.27–0.66)	78	0.52(0.43–0.61)	0.48 (0.19–0.76)	64	0.46 (0.29–0.62)	0.31 (0.11–0.51)	59
11	0.48 (0.31–0.65)	0.36(0.14–0.57)	74	0.51(0.42–0.60)	0.30 (0.10–0.50)	63	0.60 (0.45–0.76)	0.75 (0.41–1)	70
12	0.68 (0.54–0.82)	0.25 (0–0.58)	84	0.77(0.69–0.86)	0.43 (0.07–0.78)	83	0.57 (0.41–0.73)	0.63 (0.08–1)	64

PABAK=Prevalence- and bias-adjusted kappa

K_w=Quadratic-weighted kappa coefficient

CI= Confidence intervals. For items 1,3,4,6,7, and 11 for K_w in section 2, and for items 2 and 10 for K_w in section 3 confidence intervals were calculated manually using the standard error calculated in the software program STATA.