

Measuring Patient Safety in Primary Care: The Development and Validation of the "Patient Reported Experiences and Outcomes of Safety in Primary Care" (PREOS-PC)

Ignacio Ricci-Cabello, PhD¹

Anthony J. Avery, PhD²

David Reeves, PhD³

Umesh T. Kadam, PhD^{4,5}

Jose M. Valderas, PhD⁶

¹Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom

²Division of Primary Care, School of Medicine, University of Nottingham Medical School, Queen's Medical Centre, Nottingham, United Kingdom

³Centre for Biostatistics, Institute of Population Health, University of Manchester, Manchester, United Kingdom

⁴Arthritis Primary Care Research Centre, Institute for Primary Care and Health Sciences, University of Keele, Keele, United Kingdom

⁵Health Services Research Unit, Institute for Science and Technology in Medicine, University of Keele, Keele, United Kingdom

⁶Health Services & Policy Research Group, Patient Centred Care, University of Exeter Collaboration for Academic Primary Care (APEX), University of Exeter Medical School, University of Exeter, Exeter, United Kingdom

ABSTRACT

PURPOSE We set out to develop and validate a patient-reported instrument for measuring experiences and outcomes related to patient safety in primary care.

METHOD The instrument was developed in a multistage process supported by an international expert panel and informed by a systematic review of instruments, a meta-synthesis of qualitative studies, 4 patient focus groups, 18 cognitive interviews, and a pilot study. The trial version of Patient Reported Experiences and Outcomes of Safety in Primary Care (PREOS-PC) covered 5 domains and 11 scales: practice activation (1 scale); patient activation (1 scale); experiences of patient safety events (1 scale); harm (6 scales); and general perceptions of patient safety (2 scales). The questionnaire was posted to 6,736 patients in 45 practices across England. We used "gold standard" psychometric methods to evaluate its acceptability, reliability, structural and construct validity, and ability to discriminate among practices.

RESULTS 1,244 completed questionnaires (18.5%) were returned. Median item-specific response rate was 91.3% (interquartile range 28.0%). No major ceiling or floor effects were observed. All 6 multi-item scales showed high internal consistency (Cronbach's α 0.75-0.96). Factor analysis, correlation between scales, and known group analyses generally supported structural and construct validity. The scales demonstrated a heterogeneous ability to discriminate between practices. The final version of PREOS-PC consisted of 5 domains, 8 scales, and 58 items.

CONCLUSIONS PREOS-PC is a new multi-dimensional patient safety instrument for primary care developed with experts and patients. Initial testing shows its potential for use in primary care, and future developments will further address its use in actual clinical practice.

Ann Fam Med 2016;14:253-261. doi: 10.1370/afm.1935.

INTRODUCTION

Patient safety, defined by the World Health Organization as "the prevention of errors and adverse effects to patients associated with health care,"¹ is a growing interest in primary care systems.² Despite the potential impact on population health, major gaps remain in our understanding of primary care patient safety, particularly due to the lack of appropriate measurement methods,² which limits our ability to obtain reliable and repeatable rates of events for safety improvement and for research to identify fundamental underlying causes and mechanisms.

Current tools rely almost exclusively on information supplied by health care providers (eg, safety culture questionnaires and voluntary reporting of safety events).³ A growing body of evidence, however, suggests that patients are sensitive to and able to recognize a range of problems in health care delivery^{4,5} that are not identified by traditional systems of health care monitoring.^{6,7} Patient reports constitute a reliable source of information^{8,9} and have potential to improve the systematic detection of problems in health care.¹⁰⁻¹³



Conflicts of interest: authors report none.

CORRESPONDING AUTHOR

Jose M. Valderas, PhD
University of Exeter Medical School
Smeall Building Luke's Campus
EXETER, United Kingdom EX1 2LU
J.M.Valderas@exeter.ac.uk.

Our recent systematic review of primary care patient-reported safety measures showed that such instruments largely focus on a small number of relevant dimensions, mostly related to medication problems, and do not allow for a comprehensive assessment of care safety.¹⁴

We aimed therefore to develop a patient-reported instrument for comprehensively measuring experiences and outcomes of patient safety in primary care, and to test its psychometric properties.

METHODS

Based on quality standards for instrument development and evaluation,¹⁵ these steps were followed in the development of the new measure: (1) developing the framework for questionnaire domains based on the literature and expert consensus; (2) identifying and piloting relevant domains and items; and (3) psychometric testing for characteristics including acceptability, internal consistency, construct validity, and response bias.

Conceptual Framework

Two members of the research team, supported by 2 external experts (see Acknowledgments), reviewed and discussed the conceptual models proposed for patient safety in primary care.^{1,16-22} Consensus emerged on 3 necessary elements for patients' safety events: (1) patient interaction with the health care system, including self-management; (2) standards of care (with failure to adhere to them possibly due to error, but also due to other causes); and (3) actual or potential harm to patients, conceptualized as deterioration in health, including physical, mental, and social well-being. An event was hence defined as "harm or potential harm to 1 or more patients due either to an interaction with the health care system that fails to adhere to accepted standards of care (ie, that is affected by error or systemic dysfunction), or to the intrinsic risks of health care interventions."

We extracted domains from a meta-synthesis of qualitative studies on patients' experiences and perceptions of patient safety in general practices: factors contributing to safety events, experiences of safety events (active failures and harm), and patient and provider responses to safety events.²³ Additional domains and themes were obtained from 4 focus groups with 27 primary care users,²⁴ and from 23 instruments identified in our previous systematic review.¹⁴

After removing redundant domains and combining overlapping ones, 5 main domains emerged: practice activation (what does the practice do to create a safe environment); patient activation (how proactive is the patient in relation to his or her safety); experiences

of patient safety events (errors); outcomes of patient safety events (harm); and overall perceptions of patient safety (how safe patients perceive their practice to be).

Item Identification and Instrument Refinement

An expert committee composed of 5 international experts in patient safety in primary care, 3 local experts, and 2 members of the public (see Acknowledgments) was convened to support the development of the questionnaire (Figure 1).

Items were extracted from previous instruments¹⁴ to generate an item pool, which was further populated with items proposed by the development team based on the literature reviews and the focus groups. Response scales were homogenized wherever feasible. A first draft of the questionnaire was produced and then revised in an iterative process (4 iterations over 12 months) informed by repeated feedback from the expert committee.

Four waves of cognitive testing using the think-aloud technique were undertaken, including 13 individual interviews lasting 45 to 60 minutes carried out with members of the public purposefully selected to represent a range of sociodemographic backgrounds.²⁵

In a pilot with 1,975 patients in 26 English general practices, the feasibility of administration of a pre-trial version of the instrument, The Patient Reported Experiences and Outcomes of Safety in Primary Care (PREOS-PC 0.1) was tested, and the information was also used in an additional round of expert committee feedback and 5 additional cognitive interviews.

Psychometric Evaluation

In June 2014, the trial version of the questionnaire was sent to 6,736 patients registered in 45 practices purposefully sampled to ensure maximal variation in practice size and levels of deprivation and distributed across 5 regions in the North, Center, and South of England. Each practice sent the questionnaire to a computer-generated random sample of 150 patients aged 18 years and older who had had at least 1 contact with the practice in the last 12 months. Due to funding constraints, a reminder was feasible only for 10 practices, and it was sent after an interval of approximately 2 weeks.

Information on practice characteristics is available in Supplemental Appendix 1 (<http://www.annfammed.org/content/14/3/253/suppl/DC1>). Practices were asked to complete the tool PC SafeQuest,²⁶ a measure of health care professionals' perceptions of the safety climate of their practice. Ethical approval was granted by Nottingham Research Ethics Committee (Reference 13/EM/0258; July 2013).

The acceptability of the questionnaire was evaluated through examination of individual item response

rates. Scale scores were calculated as the percentage of the maximum score achievable on all items, with scores ranging from 0 (very dissatisfied, totally dis-

agree, etc) to 100 (very satisfied, totally agree, etc). Where responses were missing for 50% or more of the items in a scale, it was scored as missing; otherwise a score was derived using the available items without any imputation.

Internal consistency was deemed acceptable where inter-item correlation coefficients were at least 0.3²⁷ and Cronbach's α at least 0.7.²⁸ Test-retest reliability was analyzed using 1-way random-effects intra-class correlations (ICC), with a threshold ICC of at least 0.7, using data from a sample of 235 respondents who had been invited to complete the instrument twice approximately 2 weeks apart.¹⁵

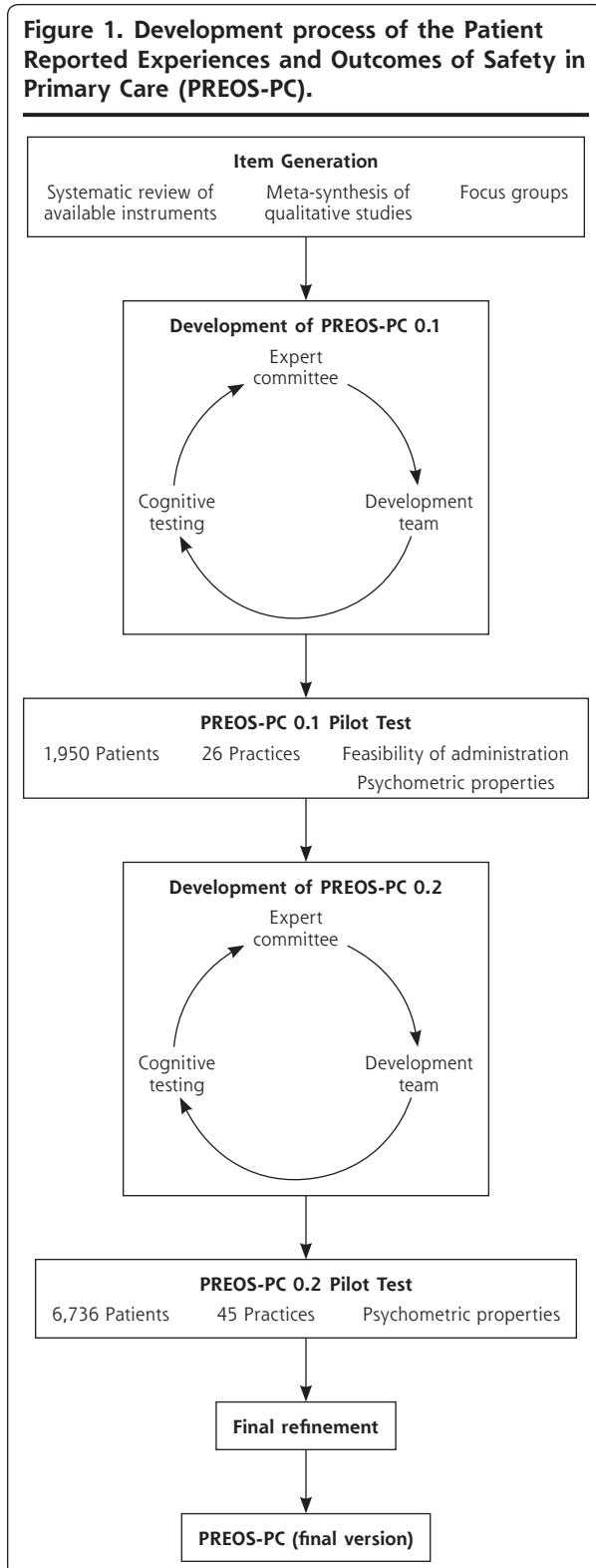
Confirmatory factor analysis was conducted to examine the construct validity of the pre-hypothesized scales. Goodness-of-fit statistics examined included the Satorra-Bentler χ^2 statistic, comparative fit index (CFI), and standardized root-mean residual (SRMR). We used Hu and Bentler's recommendation for model evaluation,²⁹ consisting in the use of a combinational rule CFI greater than 0.95 and SRMR less than 0.09. Construct validity was further examined by means of (1) pre-specified group differences, testing whether mean scores discriminated among defined groups of (a) users in line with hypothesized differences (age, ethnicity, language, country of origin, number of long-term conditions and of medications) and (b) practices (practice size, deprivation, proportion of patients aged at least 65 years, and safety climate as characterized by PC-SafeQuest); and (2) observed correlations among PREOS-PC scales with a priori hypothesized relationships.

To examine the performance of each scale as a measure of safety at the practice level, we calculated the standard error of a practice mean score as a measure of precision of measurement and the reliability coefficient (based on the between-practice intra-cluster correlation coefficient) as a measure of ability to discriminate between practices. Both measures are influenced by sample size: we based them on the mean number of patient per practice, but also estimated the sample size required to achieve reliable discrimination between practice scores at the 0.7 level.

Finally, post-hoc sensitivity analyses were carried out to examine the magnitude of potential response bias. In the subgroup of practices where reminders were sent, we used hierarchical regression models (adjusting for clustering effect) to compare patient characteristics and scale scores between patients responding to initial invitations and those responding to reminders. In order to account for skewed score distributions, bootstrap methods (50 samples) were used.

All data manipulation and analysis was conducted using STATA version 12.0 (StataCorp LP).

Figure 1. Development process of the Patient Reported Experiences and Outcomes of Safety in Primary Care (PREOS-PC).



RESULTS PREOS-PC

The Patient Reported Experiences and Outcomes of Safety in Primary Care (PREOS-PC), invites patients to report on their perceptions and experiences concerning the safety of the health care received in their primary care practice over the past 12 months (Table 1). The trial version (PREOS-PC 0.2) contained 54 standardized items and 7 open-ended questions. Forty-two standardized items were distributed across 11 scales covering all 5 domains. The remaining 12 standardized items captured details on a specific event (where did the event occur; what actions were taken, etc) and therefore were not part of any scale since their purpose was descriptive rather than evaluative.

Response Rate

The overall response rate was 18.5% (1,244/6,736), an average of 28 responses per practice. The response rate for patients who received a reminder (29.6%; 354/1,195) almost doubled that of patients who did not receive it (16.1%; 890/5541).

Compared with the overall characteristics of all eligible patients registered in the 45 participating practices, respondents were more likely to be female (59% vs 51%), at least 65 years old (39% vs 19%), and of white ethnicity (91% vs 82%) (Table 2). In our sensitivity analyses comparing demographic characteristics

and scale scores between patients responding to initial invitations and those responding to reminders, we observed that the youngest and oldest age groups and those taking less than 4 medications were less likely to respond to the first mailing (Supplemental Appendix 2, <http://www.annfammed.org/content/14/3/253/suppl/DC1>). No differences in scores between those 2 groups were observed for any of the scales, however.

Acceptability

Median item response rate was 91.3% (interquartile range 69.6% to 92.4%). When items were ranked according to nonresponse, all items in the lowest quartile pertained to the "experiences of the most recent safety problem" construct.

There was no evidence of significant ceiling or floor effects except for 2 items: "harm causing increased personal needs" and "harm causing increased financial needs" (80.1% and 80.4% of patients reporting "not at all," respectively).

Reliability

The 6 pre-hypothesized multi-item scales demonstrated high internal consistency (Cronbach's α , 0.75 to 0.96) and adequate homogeneity (inter-item correlations, 0.22 to 0.83) (Table 3). Test-retest intra-class correlation coefficients, however, were above the 0.7 standard for only 2 of the 11 scales (practice activation and harm specific to the health domain).

Practice-Level Precision and Discrimination

Taking a standard error of 5 points on the scale of 0 to 100 as indicating good precision, practice mean scores for all the globally applicable scales except patient activation demonstrated high precision. Practice means on the subset of specific scales (ie, patients who reported harm), however, showed very low precision (in all cases a standard error of more than 13 points).

Between-practice ICCs were mostly low (less than 0.03), suggesting that patient scores only weakly clustered within practices. This is reflected in the low reliability coefficients (all less than 0.7), indicating that although pre-

Table 1. Structure of the Trial Version of the Patient Reported Experiences and Outcomes of Safety in Primary Care (PREOS-PC 0.2)

Domain	Quantitative Summary		Open-Ended Questions
	Constructs	Items in Scale ^a	
Practice activation	Practice activation	11	1
Patient activation	Patient activation	2	1
Experiences of safety problems	Types of patient safety problems experienced	12	3
	Most recent experience: type of patient safety problem, location, people involved and degrees of responsibility (including patient), preventability	[12] ^b	
Outcomes of patient safety (harm)	Harm specific to health domain	5	2
	Health care, personal care, and financial needs	3	
	Time to recover from harm (type specific)	4	
	Time to recover from harm (overall)	1	
	Amount of harm experienced (overall)	1 ^c	
General perceptions of patient safety	Impact on overall health	1	
	Trustworthiness	1	0
	Overall rating of patient safety	1 ^c	
Total	12 constructs	11 scales (54 items)	7

^a Items are based on Likert scales unless otherwise noted.

^b These 12 items are not part of any scale, since their purpose is descriptive rather than evaluative.

^c Visual analog scale.

cise, the practice mean scores do not discriminate well between practices in terms of patient perceptions of safety. For most scales, however, a sample of around 100 patients would be sufficient to produce scores that discriminate well (ie, with reliability of at least 0.7).

Validity

Structural Validity

Confirmatory factor analysis was performed on the 5 multi-item scales with more than 2 items and provided evidence for high structural validity (Supplemental Appendix 3, <http://www.annfammed.org/content/14/3/253/suppl/DC1>). Three of the models met Hu and Bentler's criteria,²⁹ suggesting adequate goodness-of-fit. Moderately high item-total correlations, high internal consistency coefficients, and the results of the factor analysis indicated that each scale measures a single construct, and that the items can be combined to produce summary scores.

Table 2. Demographic Characteristics of the Participants

Characteristic	N (%)
Sex ^a	
Male	497 (41.11)
Female	712 (58.89)
Age ^b	
18-34	140 (12.03)
35-64	570 (48.97)
≥65	454 (39.00)
Ethnicity ^c	
White	1,082 (91.15)
Other ethnic group	105 (8.85)
Educational level	
Degree, degree equivalent, and above	411 (35.16)
Other qualifications	532 (45.51)
No qualifications	226 (19.33)
Health status	
Very good or good	892 (73.54)
Fair, bad, or very bad	321 (26.46)
Number of long-term conditions	
0	330 (27.99)
1	329 (27.91)
2-3	366 (31.04)
>3	154 (13.06)
Number of medications taken	
0	344 (30.10)
1-2	311 (27.21)
3-4	222 (19.42)
>4	266 (23.27)

^a Mean (SD) proportion of females registered in the 45 practices that participated in the study: 0.51 (0.05).

^b Mean (SD) proportion of eligible patients aged >65 years registered in the 45 practices that participated in the study: 0.20 (0.01).

^c Mean (SD) proportion of patients from nonwhite ethnicity registered in the 45 practices that participated in the study: 0.18 (0.04).

Construct Validity

The great majority of pairwise correlations supported our pre-specified hypothesis (Supplemental Appendix 4, <http://www.annfammed.org/content/14/3/253/suppl/DC1>). Whereas the results from the analyses of hypothesized differences between groups of patients generally supported the construct validity of the scales examined, the results from the analyses based on practice characteristics were largely inconclusive (Table 4).

Further Modifications and Final Version of PREOS-PC

Final modifications were made to PREOS-PC based on the results of the psychometric analyses (Supplemental Appendix 5, <http://www.annfammed.org/content/14/3/253/suppl/DC1>). The modifications mostly concerned the 3 single item scales in the harm domain ("time to recover from overall harm"; "amount of overall harm experienced"; and "impact of overall harm on overall health"). They were removed because they measured constructs very similar to the 3 multi-item harm-related scales that remained in the questionnaire, which demonstrated better psychometric properties. The final version of PREOS-PC includes 58 items and 8 scales (Supplemental Appendix 6, <http://www.annfammed.org/content/14/3/253/suppl/DC1>).

DISCUSSION

The PREOS-PC instrument has been developed as a tool to provide a comprehensive measure of patient-centered evaluations of patient safety in primary care, filling a gap identified in a previous systematic review.¹⁴ It was developed following the highest standards of instrument development, and this study provides preliminary evidence supporting its reliability and validity.

Strengths and Limitations

This study presents a number of methodological strengths. Evidence of the content and face validity of PREOS-PC is supported by the development of the conceptual model, the preparatory qualitative work undertaken,²³ a systematic review of instruments,¹⁴ and the iterative process of questionnaire development, which was supported by an expert committee. The questionnaire covers all of the key dimensions of our conceptual framework for primary care patient safety. It was piloted in a large sample of adults registered at a wide range of practices across England. Well-established procedures for the assessment of patient-reported instruments¹⁵ were applied to examine its reliability and validity.

In terms of limitations, our study had a low response rate (18.5%), substantially lower than response rates

Table 3. Distribution of Scores and Reliability of the Patient Reported Experiences and Outcomes of Safety in Primary Care (PREOS-PC) Scales

Domain	Construct	N	Score			Respondents With Lowest Possible Score, %	Respondents With Highest Possible Score, %
			Min	Max	Mean (SD)		
Practice activation		1,132	9.09	100	83.69 (18.01)	0	20.56
Patient activation		966	0	100	25.1 (30.8)	47.0	5.58
Experiences of safety problems		1,171	0	72.72	4.8 (9.4)	63.19	0
Outcomes of patient safety (harm)	Harm specific to the health domain	1,053	0	100	4.7 (14.1)	81.8	0.29
	Health care, personal care, and financial needs	1,043	0	91.66	2.4 (10.8)	92.6	0
	Time to recover from harm (type specific)	157	0	75	24.7 (26.7)	37.7	0
	Time to recover from harm (overall)	162	0	100	56.4 (41.6)	27.6	32.27
	Amount of harm experienced (overall)	169	0	100	35.3 (26.3)	3.0	0.60
	Impact on overall health	168	0	100	25.4 (32.5)	54.2	7.74
	Trustworthiness	1,133	0	100	87.5 (16.1)	0.3	56.78
General perceptions of patient safety	Overall rating of patient safety	1,139	0	100	86.0 (16.8)	0.2	19.79

N/A = not applicable (single item scales).

^a Based on data from 64 patients who completed the questionnaire again after 2 weeks.

^b One-way random effects intra-class correlation coefficients.

^c $P < .05$.

^d Not enough data to conduct the analyses (fewer than 10 of the respondents who completed the retest questionnaires reported harm experiences).

^e Responses needed to have sufficient cases reporting harm.

from similar large scale surveys such the GP Patient Survey,³⁰ which had a response rate of 39%. The subsample of patients who received a reminder demonstrated a substantially higher response rate (29.6%); it seems reasonable to anticipate that the inclusion of a second reminder (as was the case for the GP Patient Survey) could have increased the response rate even further.

Nonresponse can constitute a bias, since nonrespondents might differ from respondents on the key measures of interest. Meta-analyses suggest that, as long as rigorous probability sampling processes (such as those used in our study) are followed, the association between response rates and nonresponse bias within samples is generally weak.³¹ Our post-hoc analyses showed that although the low response rate resulted in an overrepresentation of elderly and polymedicated patients, this did not affect to the scale scores, suggesting that response bias did not significantly limit our estimations of the psychometric properties of the instrument.

We observed skewed score distributions for a number of items and scales. Skew is common, however, in questionnaires assessing patients' views of medical care^{32,33} and does not necessarily limit the ability to reliably distinguish practices and patient subgroups with sufficient sample sizes such as ours.³⁴

The acceptability of the "Most recent safety problem" section was relatively low, with only 60% of eligible participants adequately completing that section. This could be partially explained by potentially unclear instructions in the branching question preceding that section. This has subsequently been amended to increase clarity. It may also suggest, however, that some patients are reluctant to provide what might be considered overly detailed information about the safety problems experienced.

A substantial proportion of the scales included a low number of items, and 5 of them were based on single items. This constitutes a limitation, since short scales usually present lower levels of accuracy and reliability than scales based on higher number of items. Also, test-retest reliability could not be examined for 4 of the harm scales due to lack of sufficient cases of harm. This has minor implications for the instrument, since 3 of these have been excluded from the final version. Five of the remaining scales demonstrated low levels of test-retest reliability, suggesting that they are not adequately stable over time. This might suggest interpretation issues; further cognitive testing is needed to inform potential item modification.

Internal Consistency			Practice Mean Scores			
Cronbach's α , mean (min, max)	Inter-Item Correlation, mean (max, min)	Test-Retest Reliability ^a ICC ^b (95% CI)	Precision (standard error)	Intra-Cluster Correlation Coefficient (95%CI)	Reliability	Number of Responses Needed for 0.7 Reliability
0.89 (0.86; 0.90)	0.41 (0.39; 0.43)	0.72 (0.55-0.83) ^c	3.44	0.07 (0.03-0.11) ^c	0.66	31
0.80	0.67	0.55 (0.25-0.75) ^c	6.57	0.022 (0.001-0.050) ^c	0.33	104
0.75 (0.71; 0.76)	0.22 (0.20; 0.24)	0.57 (0.37-0.72)	1.83	0.02 (0.00-0.05)	0.36	109
0.96 (0.95; 0.96)	0.83 (0.82; 0.85)	0.72 (0.55-0.83) ^c	2.88	0.025 (0.001-0.053) ^c	0.38	90
0.88 (0.78; 0.89)	0.72 (0.63; 0.80)	-0.02 (-0.29 to 0.26)	2.23	0.019 (0.001-0.046) ^c	0.31	118
0.81 (0.71; 0.86)	0.52 (0.45; 0.67)	Not estimated ^d	13.59	0.057 (0.001-0.21)	0.18	309 ^e
N/A	N/A	Not estimated ^d	19.50	0.17 (0.001-0.34) ^c	0.45	85 ^e
N/A	N/A	Not estimated ^d	13.57	0.000 (0.000-0.126)	-	-
N/A	N/A	Not estimated ^d	15.55	0.11 (0.001-0.256) ^c	0.32	148 ^e
N/A	N/A	0.26 (-0.02 to 0.50) ^c	3.16	0.032 (0.002-0.061) ^c	0.45	71
N/A	N/A	0.24 (-0.03 to 0.48) ^c	3.29	0.029 (0.001-0.058) ^c	0.43	78

We computed scale scores for patients responding to more than 50% of scale items. Measurement errors will be somewhat larger for patients close to the 50% threshold; a stricter threshold, however, would result in more patients being fully excluded from the calculation of practice-level scores, potentially increasing the error and bias on those scores, particularly if item non-response is related to patient characteristics or experience. We considered 50% to offer a reasonable balance between these 2 sources of error and bias. Also, analyses of the psychometric properties were not stratified by levels of service use, and therefore we cannot ascertain the extent to which reliability of the scales was influenced by the number of interactions that patients had with their primary care providers.

Finally, some features of the scales are worth noting, namely the extremely high Cronbach's α in "harm specific health domains" (0.96, which may suggest item redundancy); the low inter-item correlation in the "experiences of safety problems" scale (0.22, which suggests that problems were quite independent among them); and the low test-retest coefficient for "harm: health care, personal care, and financial needs" (-0.02, presumably a result of the low number of patients reporting harm in our retest sample).

Future Steps

Further work is needed before general application of the instrument. Additional developments will include the assessment of the instrument's responsiveness to change (important if the instrument is to be used as an outcome measure in intervention studies). The development of formal methods for interpretation of the scores is pending, although provider benchmarking may in itself substantially contribute to this aim. In addition, further work comparing levels of patient safety as measured with PREOS-PC against other measures of the concept is still needed to support the validity of the instrument. Although versions of the current length may be appropriate for research purposes, shorter versions may present some advantages for service improvement. Rasch modeling is especially suitable to identify redundant items.³⁵ This work is currently underway, so is the examination of the acceptability and validity of alternative methods for administration (online and in the practice). Future steps will also include the translation of PREOS-PC into a number of different languages, and its cross-cultural adaptation and validation.

In sum, then, PREOS-PC provides a comprehensive measure of patient-reported experiences and

Table 4. Known Group Analysis Based on Characteristics of Patients and Practices

	Practice Activation ^a	Experiences of Safety Problems ^b	Impact on Health (Health Domain Specific) ^b	Impact on Health Care, Personal Care, and Financial Needs ^b	Overall Rating of Patient Safety ^a
Patient characteristics, β (95%CI)^c					
Number of long-term conditions		0.49 (0.53-0.93)	1.58 (0.82-2.35)	0.88 (0.36-1.40)	0.11 (-0.42 to 0.64)
Number of medications	0.03 (-0.34 to 0.39)	0.14 (-0.6 to 0.33)	0.65 (0.13-1.16)	0.49 (0.19-0.80)	0.17 (-0.17 to 0.53)
English as a second language					
Yes (n = 87)	1	1	1	1	1
No (n = 1,118)	5.71 (0.57-10.86)	-4.37 (-8.30 to -0.44)	-2.83 (-7.65 to 1.98)	7.87 (2.95-12.80)	-4.67 (-10.36 to 1.02)
Born in the UK					
Yes (n = 1,093)	1	1	1	1	1
No (n = 112)	-2.46 (-6.95 to 2.03)	2.83 (-0.02 to 5.68)	1.48 (-2.38 to 5.34)	-5.13 (-8.64 to -1.63)	3.13 (-0.19 to 6.47)
Practice characteristics, r (95%CI)^d					
Practice size	0.24 (-0.06 to 0.50)	-0.27 (-0.52 to 0.03)	-0.11 (-0.40 to 0.19)	-0.02 (-0.31 to 0.28)	0.01 (-0.20 to 0.39)
Proportion of patients aged >65 years	0.06 (-0.25 to 0.35)	-0.08 (-0.37 to 0.22)	-0.07 (-0.36 to 0.24)	-0.04 (-0.34 to 0.26)	0.14 (-0.16 to 0.42)
Deprivation score ^e	-0.32 (-0.56 to -0.02)	0.30 (0.00-0.55)	0.20 (-0.10 to 0.47)	0.18 (-0.13 to 0.45)	-0.36 (-0.60 to -0.07)
Safety Climate ^f	-0.09 (-0.45 to 0.29)	0.06 (-0.32 to 0.42)	0.10 (-0.28 to 0.46)	0.15 (-0.50 to 0.25)	0.07 (-0.31 to 0.43)

^a Higher scores indicate perception of safer practices.

^b Higher scores indicate more severe/frequent access problems or more severe harm.

^c Bivariate linear regression.

^d Correlation coefficient.

^e Deprivation measured using the Index of Multiple Deprivation; higher scores indicate higher deprivation levels.

^f Safety climate measured using the instrument "PC-SafeQuest" (information available only from 31 practices).

outcomes of safety in primary care. Results from psychometric analysis support its internal consistency and validity, though findings for test-retest reliability were mixed. Further work is needed before general application of the instrument.

To read or post commentaries in response to this article, see it online at <http://www.annfammed.org/content/14/3/253>.

Key words: patient safety; primary care; patient-centered care; health care evaluation mechanisms; health care surveys

Submitted August 20, 2015; submitted, revised, December 7, 2015; accepted December 21, 2015.

Funding support: This research is part-funded by the UK National Institute for Health Research School for Primary Care Research (NIHR SPCR). The views expressed are those of the authors and not necessarily those of the NIHR, the NHS, or the Department of Health.

Acknowledgments: We would like to thank the following participants and organizations involved in the development of the questionnaire:

Dr Itziar Larizgoitia (WHO), Prof Tony Avery (University of Nottingham), Prof Stephen Campbell (University of Manchester), Prof Charles Vincent

(University of Oxford), Dr Angela Coulter (University of Oxford), Dr Sarah P Slight (University of Durham), Dr Umesh Kadam (University of Keele and previously Arthritis Research UK Primary Care Centre), Ms Liz Thomas (Action against Medical Accidents (AvMA)), Mr Derek Shaw, and Mr Antony Chuter (members of the public) for their participation as members of the Expert Committee.

Dr Daniela Gonçalves Bradley and Dr Suzanne Shale (Oxford University) for their contribution in the development of the conceptual framework of patient safety.

Mr Antony Chuter (member of the public) and Ms Liz Taylor (Association against Medical Accidents (AvMA)) for their support in recruiting members of the public for the focus groups.

Ms Kate Marsden (University of Nottingham), Dr Katherine Perryman (University of Manchester), Ms Jane Barnett (University of Southampton), Dr Ian Litchfield (University of Birmingham), Ms Sally Thomas (University of Keele) and all the health professionals in the 45 practices involved in this project for distributing the questionnaires to the patients as part of its pilot-testing.

Dr Brian Bell for providing the data on the characteristics of the practices.

Finally we would like to thank all the patients and members of the public that participated in the cognitive interviews, focus groups, and completing the survey.

Supplementary materials: Available at <http://www.AnnFamMed.org/content/14/3/253/suppl/DC1/>.

References

- World Health Organization. Patient safety. <http://www.euro.who.int/en/health-topics/Health-systems/patient-safety/patient-safety>. Accessed Aug 19, 2015.
- Lorincz CY, Drazen E, Sokol PE, et al. Research in ambulatory patient safety 2000–2010: A 10-year review. https://c.ymcdn.com/sites/npsf.site-ym.com/resource/resmgr/PDF/Research-in-Amb-Pat-Saf_AMAr.pdf. Accessed Aug 19, 2015.
- Spencer R, Campbell SM. Tools for primary care patient safety: a narrative review. *BMC Fam Pract*. 2014;15:166.
- Weingart SN, Pagovich O, Sands DZ, et al. Patient-reported service quality on a medicine unit. *Int J Qual Health Care*. 2006;18(2):95-101.
- Schwappach DL. Review: engaging patients as vigilant partners in safety: a systematic review. *Med Care Res Rev*. 2010;67(2):119-148.
- Levtzion-Korach O, Frankel A, Alcalai H, et al. Integrating incident data from five reporting systems to assess patient safety: making sense of the elephant. *Jt Comm J Qual Patient Saf*. 2010;36(9):402-410.
- Evans SM, Berry JG, Smith BJ, Esterman AJ. Consumer perceptions of safety in hospitals. *BMC Public Health*. 2006;6:41.
- King A, Daniels J, Lim J, Cochrane DD, Taylor A, Ansermino JM. Time to listen: a review of methods to solicit patient reports of adverse events. *Qual Saf Health Care*. 2010;19(2):148-157.
- Bjertnaes O, Deilkås ET, Skudal KE, Iversen HH, Bjerkan AM. The association between patient-reported incidents in hospitals and estimated rates of patient harm. *Int J Qual Health Care*. 2015;27(1):26-30.
- Reader TW, Gillespie A, Roberts J. Patient complaints in healthcare systems: a systematic review and coding taxonomy. *BMJ Qual Saf*. 2014;23(8):678-689.
- Weingart SN, Pagovich O, Sands DZ, et al. What can hospitalized patients tell us about adverse events? Learning from patient-reported incidents. *J Gen Intern Med*. 2005;20(9):830-836.
- Lawton R, O'Hara JK, Sheard L, et al. Can staff and patient perspectives on hospital safety predict harm-free care? An analysis of staff and patient survey data and routinely collected outcomes. *BMJ Qual Saf*. 2015;24(6):369-376.
- Donaldson LJ. The wisdom of patients and families: ignore it at our peril. *BMJ Qual Saf*. 2015;24(10):603-604.
- Ricci-Cabello I, Gonçalves DC, Rojas-García A, Valderas JM. Measuring experiences and outcomes of patient safety in primary care: a systematic review of available instruments. *Fam Pract*. 2015;32(1):106-119.
- Valderas JM, Ferrer M, Mendivil J, et al; Scientific Committee on "Patient-Reported Outcomes" of the IRYSS Network. Development of EMPRO: a tool for the standardized assessment of patient-reported outcome measures. *Value Health*. 2008;11(4):700-708.
- Council of the European Union. European Council recommendation on patient safety, including the prevention and control of healthcare associated infections. http://ec.europa.eu/health/ph_systems/docs/patient_rec2009_en.pdf. Published Jun 5, 2009. Accessed Aug 19, 2015.
- Kohn LT, Corrigan JM, Donalson MS, Institute of Medicine. *To Err is Human: Building a Safer Health System*. Washington, DC: National Academy Press; 2000.
- Agency for Healthcare Research and Quality. PSNet Glossary. PSNet patient safety network Web site. <http://psnet.ahrq.gov/glossary.aspx>. Accessed Aug 19, 2015.
- Henriksen K, Battles JB, Keyes MA, Grady ML. *Advances in Patient Safety: New Directions and Alternative Approaches*. Rockville, MD: Agency for Healthcare Research and Quality; 2008.
- Vincent CA. *Patient Safety*. 2 ed. Oxford, England: Wiley Blackwell; 2010.
- Runciman W, Hibbert P, Thomson R, Van Der Schaaf T, Sherman H, Lewalle P. Towards an International Classification for Patient Safety: key concepts and terms. *Int J Qual Health Care*. 2009;21(1):18-26.
- The Chronic Care Model. Improving Chronic Illness Care Web site. http://www.improvingchroniccare.org/index.php?p=Model_Elements&s=18. Published 2003. Accessed Aug 19, 2015.
- Ricci-Cabello I, Gonçalves DC, Campbell S, Slight S, Valderas JM. Patients' experiences of patient safety in primary care in England: a systematic review and meta-synthesis. Presented at: North American Primary Care Research Group (NAPCRG) Annual Meeting; November 21-25, 2014; New York, NY.
- Ricci-Cabello I, Berenguer A, Pujol-Ribera E, Pons-Vigues M, Valderas JM. [Exploring patients' perceptions of patient safety in primary care in England: a qualitative study]. *Gac Sanit*. 2014;28: 256-256. <http://gacetasanitaria.org/es/pdf/X021391114410206/S300/>.
- Ericsson KA. Protocol analysis and expert thought: Concurrent verbalizations of thinking during experts' performance on representative tasks. In: Ericsson KA, Charness N, Feltovich PJ, Hoffman RR, eds. *The Cambridge Handbook of Expertise and Expert Performance*. Cambridge, England: Cambridge University Press; 2006:223-242.
- de Wet C, Spence W, Mash R, Johnson P, Bowie P. The development and psychometric evaluation of a safety climate measure for primary care. *Qual Saf Health Care*. 2010;19(6):578-584.
- Streiner DL, Norman GR. *Health Measurement Scales: A Practical Guide to Their Development and Use*. Oxford, England: Oxford University Press; 2008.
- Nunnally JC, Bernstein IH. *Psychometric Theory*. 3rd Ed. New York, NY: McGraw-Hill; 1994.
- Hu L-t, Bentler PM. Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychol Methods*. 1998;3(4):424-453.
- Campbell J, Smith P, Nissen S, Bower P, Elliott M, Roland M. The GP Patient Survey for use in primary care in the National Health Service in the UK—development and psychometric characteristics. *BMC Fam Pract*. 2009;10:57.
- Groves RM. Nonresponse Rates and Nonresponse Bias in Household Surveys. *Public Opin Q*. 2006;70(5):646-675.
- Elliott MN, Zaslavsky AM, Goldstein E, et al. Effects of survey mode, patient mix, and nonresponse on CAHPS hospital survey scores. *Health Serv Res*. 2009;44(2 Pt 1):501-518.
- Mead N, Bower P, Roland M. The General Practice Assessment Questionnaire (GPAQ) - development and psychometric characteristics. *BMC Fam Pract*. 2008;9:13.
- Elliott MN, Haviland AM, Kanouse DE, Hambarsoomian K, Hays RD. Adjusting for subgroup differences in extreme response tendency in ratings of health care: impact on disparity estimates. *Health Serv Res*. 2009;44(2 Pt 1):542-561.
- Tennant A, McKenna SP, Hagell P. Application of Rasch analysis in the development and application of quality of life instruments. *Value Health*. 2004;7(Suppl 1):S22-S26.