

Experience of Inpatient Mental Health Care Assessed With Service User–Developed and Conventional Patient-Reported Outcome Measures

Ioannis Bakolis, Ph.D., Prashant Gupta, M.D., Til Wykes, Ph.D.

Objective: The goal of this study was to examine and compare the psychometric properties of a patient-reported outcome measure (PROM) generated with patients' input (Views on Inpatient Care [VOICE]) and a PROM conventionally generated without patients' input (Service Satisfaction Scale: Residential Services Evaluation [SSS-Res]) for assessing a patient's perception of psychiatric ward care.

Methods: In a stepped-wedge cluster-randomized trial conducted in the United Kingdom, 1,058 participants admitted to 16 wards reported on their perceptions of care via VOICE and SSS-Res before or up to 2 years after the staff training. Exploratory and confirmatory factor analyses were used to investigate the structure of the PROMs and to assess reliability and convergent validity as well as sensitivity to change; the analyses also considered whether study participants had been admitted voluntarily to the ward.

Results: Two factors emerged from VOICE, labeled "trust" and "involvement," and from SSS-Res, labeled "environment" and "care," at baseline. All subscales had high internal consistency and good convergent validity. An ability to detect change in care due to the staff training was observed on the trust subscale of VOICE ($N=1,058$, mean difference = -0.25 , 95% CI = -0.48 to -0.02), but no change was detected on any of the SSS-Res subscales. Patients admitted involuntarily benefited the most from the staff training.

Conclusions: VOICE captured patients' perceptions of ward care better than SSS-Res and was sensitive to changes in aspects of trust, suggesting that participatory approaches for developing PROMs improve patients' self-reports on the care they received.

Psychiatric Services 2022; 73:1132–1139; doi: 10.1176/appi.ps.202100470

The U.K. 1990 National Health Service (NHS) and Community Care Act was the first piece of legislation that established a formal requirement for involvement of users and caregivers in service planning. Since then, several legal and policy measures have ensured that those who use services have an equal say in how these services are planned, developed, and delivered. What is known as patient and public involvement is a key requirement for research activities by many funding bodies and ethics committees.

There has been much concern about the impact of patient and public involvement on health services research, with some researchers and clinicians arguing that this impact can be shown only quantitatively (1) and with others noting that a more qualitative approach is required (2). One tactic has been to make major modifications in research methods to improve measures employed in health services research. In the United States, the Patient-Centered Outcomes Research Institute (PCORI) (3) developed a way of engaging stakeholders, such as service users, in the generation of patient-reported outcome measures (PROMs). PROMs represent an advance on conventional methods, but stakeholder

engagement typically is only partial. By contrast, the Service User Research Enterprise in the United Kingdom developed a method for generating PROMs entirely from the ground

HIGHLIGHTS

- This is the first study to directly compare two patient-reported outcome measures (PROMs) assessing patient perceptions of ward care, one that was developed entirely by service users and the other consisting of a conventional measure.
- Results from the patient-generated PROM (called Views on Inpatient Care) indicated that changes in ward care were appreciated by patients, especially those who had been admitted involuntarily, whereas the conventional measure did not detect any changes in patients' perception of care.
- This study highlights the importance of developing self-report–based outcome measures sensitive to changes in service provision and encourages new participatory approaches for developing PROMs.

up, starting with assessment of people's experience with the service and with focus groups comprising patients who assess their experience with health care services. The measure development proceeds gradually to ensure that the collective experience drives the development, with psychometric testing of the measure brought in as a check only (4). Although represented as a methodological change, this model for measure development also represents a shift in whose voice is prioritized during development (5). The question remains, however, of how PROMs affect health services research. Do outcome measures generated entirely by patients who are using mental health services perform any better (or worse) than those generated without service users' input?

The aim of this study was to examine and compare the properties of a patient-generated PROM and a PROM conventionally generated without patients' input for assessing their perceptions of psychiatric ward care. More important, we were interested in the ability of these PROMs to detect changes in perception of care after service changes were implemented on the ward after staff training. We also investigated the effects of these changes on service users admitted involuntarily, that is, under a legal sanction. Patients admitted involuntarily under a legal sanction cannot leave the hospital without permission, and involuntary admission is associated with low patient satisfaction with care (6). Such patients are less likely to view ward care positively, and we therefore were particularly interested in how these patients were affected by the ward changes (7).

The study obtained data from a previously published stepped-wedge cluster-randomized trial (SW-CRT) that investigated patient perceptions of ward care after staff training to support ward-based therapeutic activity (8). The pathway from intervention to impact on patient perception of care is complex and potentially includes improvements in staff morale, changes in ward activities, provision of opportunities for patients to attend, and direct effects on patients. This study assumed that, via any route, the staff training would generally improve the therapeutic environment and that therefore these data would provide an ideal opportunity to test whether either PROM would detect an improvement in care from the patient's point of view.

METHODS

Study Design

This study is based on a secondary analysis of data derived from a cross-sectional SW-CRT (8), a type of cluster-randomized trial where the timing of the intervention is randomized such that wards randomized to receive staff training subsequently remained in the intervention arm. Two wards received staff training at a time, until all the wards had received the training (see glossary, Table S1 in the online supplement). Wards were sampled three, five, or seven times and could provide data until all staff had

received the training. Ethical approval for this study was granted by Bexley and Greenwich Research Ethics Committee (ref 07/H0809/49).

Participants

All participants provided self-report data on perceptions of care once in the period, either before or after the staff training. Participants were unaware of the trial arm in which they were included, that is, they did not know whether the ward received staff training, so the service users who received the assessments were blind to intervention allocation. All participants entered the data set once only, even if they were readmitted during the study, and therefore provided only one set of data in either the pre- or postintervention period. Patients were eligible for study participation if they could communicate in English, had been on the ward for at least 7 days, and could provide informed consent. The only exclusion criterion was previous participation in the trial. We endeavored to recruit 50% of all eligible patients at the time of data collection. This study was carried out in distinct demographic areas (see panel S1 in the online supplement). A more detailed description of the study is provided elsewhere (8).

Intervention

Sixteen wards from the NHS were provided with training in five evidence-based and feasible interventions from a menu of eight, according to guidelines and ward team judgments. A psychologist trained ward staff to deliver all the interventions. The training offered to all wards included social cognition and interaction training (9), cognitive-behavioral therapy-based communications training for nurses (cofacilitated by a service user educator), and computerized cognitive remediation therapy (to involve occupational therapists) (10); pharmacists were recruited to run medication education groups (11). According to individual ward needs, ward staff could choose more sessions from the hearing voices group (12), emotional coping skills group (13), problem-solving skills group (14), and relaxation and sleep hygiene and coping with stigma group (15). A more detailed description of the five evidence-based and feasible interventions is provided elsewhere (8). Details of the staff training can be also found on the study website (<http://www.perceive.iop.kcl.ac.uk>).

Sample Characteristics

Data were available from 1,108 participants (70% of the population eligible to participate [$N=1,583$]) who took part either before or after the staff training in the interventions (a CONSORT diagram of the main trial is available in Wykes et al. [8]). Data sufficient for the analyses were provided from 1,058 participants (96%). These participants provided a blind self-report of perceptions of care on two instruments for assessing the perceptions of service users of the care they received, Views on Inpatient Care (VOICE) and Service Satisfaction Scale: Residential Services Evaluation (SSS-Res) (see details on these two instruments below), before or up to

TABLE 1. Demographic characteristics of the patients and VOICE and SSS-Res total scores, overall and by pre- and postintervention wards

Characteristic	Overall (N=1,108)		Preintervention wards (N=670)		Postintervention wards (N=438)	
	N	%	N	%	N	%
Gender						
Men	609	55	352	53	257	59
Women	499	45	318	47	181	41
Age in years (median, IQR)	40	13	40	13	40	13
First language						
English	879	80	525	79	354	82
Not English	219	20	141	21	78	18
Race-ethnicity						
White	556	50	325	49	231	53
Mixed	71	6	38	6	33	8
Asian	59	5	30	5	29	7
Black	377	34	250	37	127	29
Chinese	3	<1	3	<1	0	—
Other ^a	41	4	23	3	18	4
Legal status of admission						
Involuntary	616	56	386	58	230	53
Voluntary	485	44	280	42	205	47
Previous admissions						
No	281	26	173	26	108	26
Yes	797	74	482	74	315	75
N of previous admissions (median, IQR)	4	6	4	6	4	5
Primary diagnosis						
Drug related	62	6	40	6	22	5
Psychosis or bipolar disorder	677	63	423	64	254	62
Depression or anxiety	132	12	80	12	52	13
Other	196	18	116	18	80	20
Length of stay in days (median, IQR)	97	222	111	246	64	151
Measured clinical outcomes						
VOICE total scores (median, IQR)	56	18	57	19	54	17
SSS-Res total scores (median, IQR)	89	26	91	27	86	24

^a Other included Arab, Sikh, and Jewish. IQR, interquartile range; SSS-Res, Service Satisfaction Scale: Residential Services Evaluation; VOICE, Views on Inpatient Care.

2 years after staff training between November 2008 and January 2013.

Statistical Power and Sample Size

A sample size of 1,058 in a standard cluster-randomized design would have given approximately 90% power to detect a standardized effect size of 0.5 (moderate), using double-sided significance tests with $\alpha=0.05$. Because of the stepped-wedge design, the actual number of wards and participants in the intervention and control groups varied among time points, so the power and sample size calculations were approximate but were designed to be conservative.

Measures

VOICE. VOICE (7) is a 19-item multifaceted self-report measure developed by service users via participatory methods

and has good feasibility and psychometric properties. It was developed iteratively through an innovative participatory methodology to maximize service user involvement. The development followed several stages. A topic guide was developed through a literature search guided by a reference patient group. Repeated focus groups of service users were then convened to generate qualitative data. One of the groups specifically included participants who had been detained under the Mental Health Act (1983) because we anticipated that their care experience may differ from that of patients receiving care voluntarily. The data were thematically analyzed by service researchers, who then generated a draft measure that was refined by expert panels of users and the reference user patient group. VOICE assesses service users' perception of acute care in relation to trust and respect, including items such as "I was made to feel welcome when I arrived on this ward" as well as items on therapeutic contact and care. The key score was the total, ranging from 19 to 114, with higher scores indicating a worse perception of care.

SSS-Res. SSS-Res (16) is a 33-item measurement instrument used in previous studies of inpatient care to assess client satisfaction in mental health and other human service settings, including items such as "Knowledge and competence of staff seen." The key outcome was the total score, which ranged from 33 to 165, with higher scores indicating worse satisfaction with care.

Background information. This information included age, gender, race-ethnicity, primary diagnosis, first language, length of stay (up to entry into the study), and whether patients were detained involuntarily (i.e., under a legal sanction).

Statistical Analysis

Descriptive statistics were calculated for all included measures. All analyses were conducted with Stata, version 15.1.

Exploratory factor analysis. Apart from the obvious differences in item generation, we wanted to understand the make-up of the items and whether VOICE and SSS-Res indexed the same underlying constructs. We therefore used an exploratory factor analysis (EFA) of the polychoric correlation matrix of the 19 VOICE and 33 SSS-Res items to determine and confirm scale factor structures on the data

collected before any intervention (N=670). A varimax rotation was applied to improve the interpretability of the factors. Three criteria were used to select the final factors: a scree plot, eigenvalues >1, and >90% of total variance explained by the factors.

Psychometric evaluation. Scaling assumptions were investigated by using a confirmatory factor analysis (CFA) model fitted to data collected after the intervention (N=438). CFA was applied by using the weighted least-squares estimator with a mean- and variance-adjusted chi-square method to handle ordered categorical items (17). Missing data for both measures were handled by using full-information maximum likelihood estimation. This method computes parameter estimates on the basis of all available data, including incomplete cases (i.e., assuming that data are missing at random). To evaluate the overall model fit, the comparative fit index (CFI) (18), the Tucker-Lewis index (TLI) (19), and the root mean square error of approximation (RMSEA) (20) were calculated. CFI and TLI values of >0.90 indicate adequate fit (21). An RMSEA value of <0.05 indicates close fit (21), between 0.05 and 0.09 suggests adequate fit, and ≥ 0.10 suggests poor fit.

Reliability was assessed by examining the scale internal consistency with Cronbach's alpha, with an $\alpha > 0.70$ indicating appropriate internal consistency for each subscale for data collected after the intervention (22). Convergent validity assesses the ability of the PROM instrument to yield consistent, reproducible estimates by assessing hypothesized relationships with similar constructs. Convergent validity was examined by estimating the correlation between VOICE and the SSS-Res dimensions with Spearman's correlation coefficients for data collected after the intervention. These correlations were interpreted as follows: >0.90, excellent relationship; 0.71–0.90, good relationship; 0.51–0.70, fair relationship; 0.31–0.50, weak relationship; and ≤ 0.30 , no relationship. Correlations of the total scores of VOICE and SSS-Res were also calculated.

Ability to detect a change considers whether the instrument can identify differences in scores over time among individuals or groups whose perspectives have changed with respect to the measurement concept. The approach adopted in this study differed from the previous approach (8) by following new analytic developments for stepped-wedge designs (23–25). It adopted the guidance for a cross-sectional SW-CRT with 16 clusters. The approach used generalized linear mixed models with a jackknife procedure and included as dependent variable each standardized total and derived factor scores for both measures; intervention

TABLE 2. Factor loadings for the two factors of VOICE for the preintervention sample (N=670)^a

Items	VOICE-T	VOICE-I
I was made to feel welcome when I arrived on this ward	.42	.41
I have a say in my care and treatment	.25	.63 ^b
Ward rounds are useful for me	.15	.63 ^b
I feel my medication helps me	.17	.55 ^b
I have the opportunity to discuss my medication and side effects	.17	.69 ^b
Staff give me medication instead of talking to me	.01 ^c	.41
Staff take an interest in me	.41	.47
Staff are available to talk to when I need them	.47	.43
I trust the staff to do a good job	.63 ^b	.35
I feel that staff understand how my illness affects me	.51 ^b	.45
I feel that staff treat me with respect	.64 ^b	.28
I think the activities on the ward meet my needs	.46	.19
I find one-to-one time with staff useful	.53 ^b	.19
I find it easy to keep in contact with family and friends	.58 ^b	-.13 ^c
I feel safe on the ward	.49	.14
I feel staff respond well when the panic alarm goes off	.37	.20
I feel staff respond well when I tell them I am in crisis	.57 ^b	.20
I feel able to practice my religion while I am in hospital	.49	.22
I think staff respect my ethnic background	.59 ^b	.31

^a Correlation coefficients between each of the VOICE items and trust and involvement subscales; larger values indicate higher correlation between the VOICE items and trust and involvement subscales. Cronbach's alphas are 0.85 and 0.78 for VOICE-T and VOICE-I, respectively. VOICE-I, Views on Inpatient Care Involvement; VOICE-T, Views on Inpatient Care Trust.

^b Item with >0.5 factor loadings.

^c Item with <0.1 factor loadings.

was an independent binary variable, time was a categorical variable, and gender and ward were fixed effects for both the standardized total and derived factor scores for both measures. We also examined whether legal sanction status of a patient modified the ability to detect changes in standardized total and derived factor scores for both measures due to the intervention.

RESULTS

In total, 670 service users provided data before the staff training and 438 after the training. Participant characteristics did not differ between the pre- and postintervention samples. The sociodemographic characteristics of the patients and the scores on the two PROMs are presented in Table 1.

Exploratory Factor Analysis

We retained two factors for both VOICE and SSS-Res scales at baseline in the EFA because they explained the largest percentage (>95%) of total variance, had eigenvalues >1, and allowed a meaningful interpretation. The two VOICE factors related to trust (VOICE-T) and involvement (VOICE-I) (Table 2), and the SSS-Res factors related to environment (SSS-Res-E) and care (SSS-Res-C) (Table 3).

Psychometric Evaluation

Scaling assumptions. The two-factor CFA models had relatively good fit for both VOICE and SSS-Res scales. An RMSEA value of <0.05 and CFI and TLI values of >0.9 suggested adequate fit (see also Figures 1 and 2).

TABLE 3. Factor loadings for the two factors of SSS-Res for the preintervention sample (N=670)^a

Items	SSS-Res-E	SSS-Res-C
Kinds of services offered	.64 ^b	.25
Opportunity to choose which staff you see	.40	.37
How much services helped you deal with your problems	.48	.44
Office procedures (scheduling, forms, tests, others)	.58 ^b	.24
The kinds of questions asked and how they were asked	.52 ^b	.33
Knowledge and competence of staff seen	.69 ^b	.19
Location and access to services (distance, ease of parking)	.47	.20
Appearance and layout of the facility and grounds	.66 ^b	-.03 ^c
Ability of staff you worked with to listen and understand	.57 ^b	.36
Personal manner, involvement, and caring of the staff	.65 ^b	.29
Program activities (Alcoholics Anonymous, emotions, recovery, others)	.46	.26
Cleanliness and comfort of the residential environment	.65	.05 ^c
How your family or other support people received support	.42	.32
Help with practical problems (financial, locating housing, other challenges)	.21	.49
Effect of services in helping you stay well and prevention	.27	.58 ^b
Confidentiality and respect for your rights as an individual	.47	.46
Amount of help you have received	.49	.47
Information on how to get the most out of services available	.32	.59 ^b
Helping you get needed prescriptions or other medical assistance	.17	.65 ^b
Helping you handle medication side effects, discomfort, or other problems	.25	.62 ^b
Suggestions on what to do on your own after discharge	.02 ^c	.78 ^b
Explanations of agency procedures and treatment plans	.17	.72 ^b
Effect of services in helping relieve symptoms	.24	.66 ^b
Response of staff to your urgent needs during day	.59 ^b	.36
Response of staff to your urgent needs during evening/night	.47	.41
Safety of the environment (how "at home" you could be)	.54 ^b	.26
Usefulness of referrals to other counselors, doctors	.17	.69 ^b
Communication between residential staff and other services	.33	.56 ^b
Willingness to see you as often as you feel is needed	.47	.46
Handling and accuracy of your records	.53 ^b	.31
Quality and quantity of food	.46	.19
Help you received from working on problems with others	.40	.37
In an overall general sense, how satisfied are you with services?	.56 ^b	.45

^a Correlation coefficient values between each of the SSS-Res items and environment and care subscales; larger values indicate higher correlation between the SSS-Res items and environment and care subscales. Cronbach's alphas are 0.93 and 0.92 for SSS-Res-E and SSS-Res-C, respectively. SSS-Res-C, Service Satisfaction Scale: Residential Services Evaluation-Care; SSS-Res-E, Service Satisfaction Scale: Residential Services Evaluation Environment.

^b Item with >0.5 factor loadings.

^c Item with <0.1 factor loadings.

Reliability and convergent validity. VOICE-T, VOICE-I, SSS-Res-E, and SSS-Res-C were satisfactorily reliable, with Cronbach's alphas of 0.85, 0.78, 0.93, and 0.92, respectively (Tables 2 and 3). The total scores and factor scores were correlated between VOICE and SSS-Res before and after the intervention (Pearson correlation coefficients >0.7 and >0.9; see Table S2 in the online supplement).

Ability to detect change. The ability to detect a change in patients' perception of care after the intervention was evident for the total VOICE score (mean difference [MD]= -0.29, 95% confidence interval [CI]= -0.54 to -0.05, N=1,058) and for the VOICE-T factor score (MD= -0.25, 95% CI= -0.48 to -0.02, N=1,058). The VOICE-I scale (MD= -0.15, 95% CI= -0.42 to 0.11, N=1,058) and

none of the SSS-Res measures captured this effect (total scores of SSS-Res, MD= -0.24, 95% CI= -0.52 to 0.15, N=1,025; of SSS-Res-E, MD= -0.18, 95% CI= -0.49 to 0.12, N=1,025; and of SSS-Res-C, MD= -0.14, 95% CI= -0.48 to 0.18, N=1,025).

Legal sanction status at admission modified the effect of the intervention on the VOICE total score (p for interaction=0.023) and factor score of the VOICE-T scale (p for interaction=0.031) but not on the VOICE-I scale (p for interaction=0.504). The intervention had a significant effect on patients' perception of care among those who received inpatient care after involuntary admissions for both the total VOICE score (MD= -0.48, 95% CI= -0.88 to -0.08, N=582) and VOICE-T score (MD= -0.38, 95% CI= -0.016 to -0.08, N=582). Among patients admitted voluntarily, no evidence was found for an intervention effect on either the total VOICE score (MD= -0.01, 95% CI= -0.23 to 0.22, N=469) or VOICE-T score (MD= -0.03, 95% CI= -0.45 to 0.39, N=469). By comparison, evidence was detected for a significant interaction effect between the intervention and legal sanction at admission for the SSS-Res-E score (p=0.002) but not for total SSS-Res score (p=0.100) or SSS-Res-C (p=0.998). Again, people admitted involuntarily reported benefits af-

ter the intervention on the SSS-Res-E factor score (MD= -0.36, 95% CI= -0.68 to -0.04, N=566), but no significant effect on this score was detected for those admitted voluntarily (MD= -0.03, 95% CI= -0.45 to 0.39, N=459).

DISCUSSION

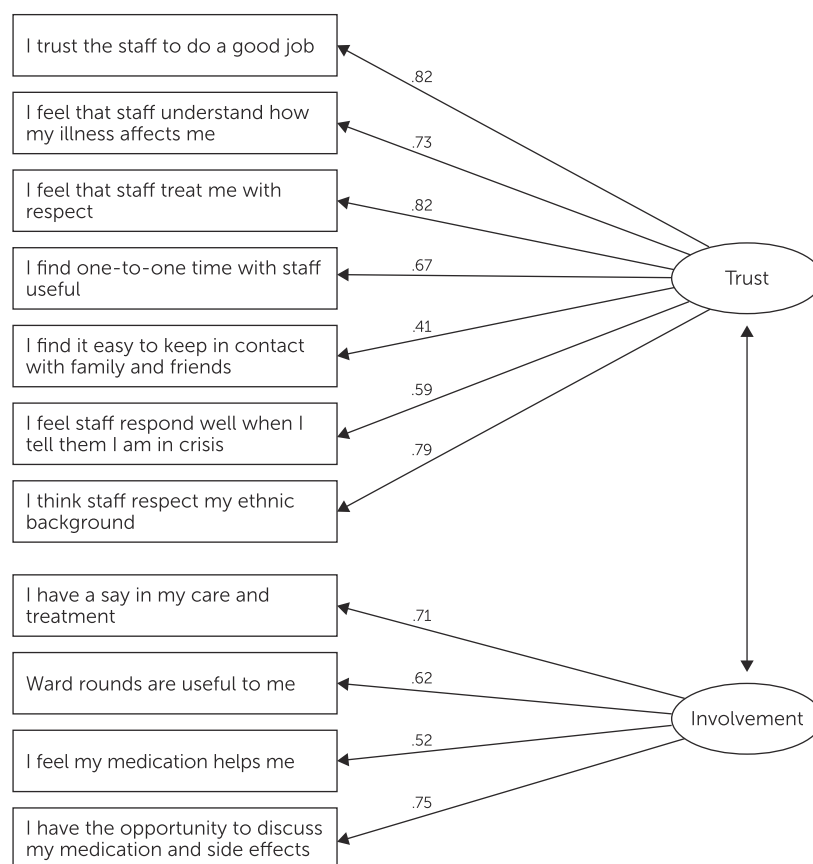
This is the first study to compare data from PROMs that were developed and generated either by service users or without such input. Both measures had subscales and exhibited similarly satisfactory reliability and validity. However, we identified differences in their ability to detect improvements in patients' perception of mental health care following staff training in ward-based therapeutic activity, with strong evidence of an effect of the staff training

intervention on the measures assessed with the patient-generated scale, VOICE. We also observed a more pronounced benefit of the intervention for an important target group—patients admitted involuntarily, that is, under legal sanction. Although we found no overall effect for the intervention when using the conventionally generated scale, SSS-Res, we observed an effect when using the SSS-Res-E subscale among those who were involuntarily admitted. Findings from a recent study indicate that individuals in the United Kingdom are less satisfied with psychiatric inpatient care compared with individuals in other countries, and individuals who were admitted involuntarily had the strongest association with dissatisfaction scores (26). This finding emphasizes that changes in therapeutic activities may have the strongest effect for those who are the least satisfied with services. We conclude that the two instruments differ in their ability to detect change in patients' perception of care, with the more conventionally derived measure, SSS-Res, not revealing many differences. This difference between the two instruments existed despite both scales having more than adequate psychometric properties and being highly correlated with each other.

To our knowledge, this is the largest sample in which psychometric analysis was conducted with the VOICE measure. Although the total score of VOICE and the factor scores of the VOICE-T and VOICE-I subscales highly correlated with the SSS-Res total score and the factor scores of SSS-Res-E and SSS-Res-C, we noted some distinct differences between the two measures. Issues of trust and involvement were given more weight during the development of the VOICE instrument, and items on diversity were included in this instrument that did not appear in SSS-Res. Conversely, items regarding the physical environment and office procedures feature in the SSS-Res, but participants involved in VOICE development did not consider these items as important as other items, and therefore they were not included. It is impossible to accurately assess inpatient care without involving the people directly affected by that form of care. Developing a real PROM, an outcome instrument that by definition is valued by patients who are using the services, is essential in any evaluation and development of inpatient services.

The inpatient wards selected in our study served inner-city and suburban populations of different socioeconomic backgrounds and provided care for individuals with a variety of diagnoses, comparable to many other inpatient wards in

FIGURE 1. Confirmatory factor analysis of items on the trust and involvement subscales of VOICE for the postintervention sample (N=438 patients)^a

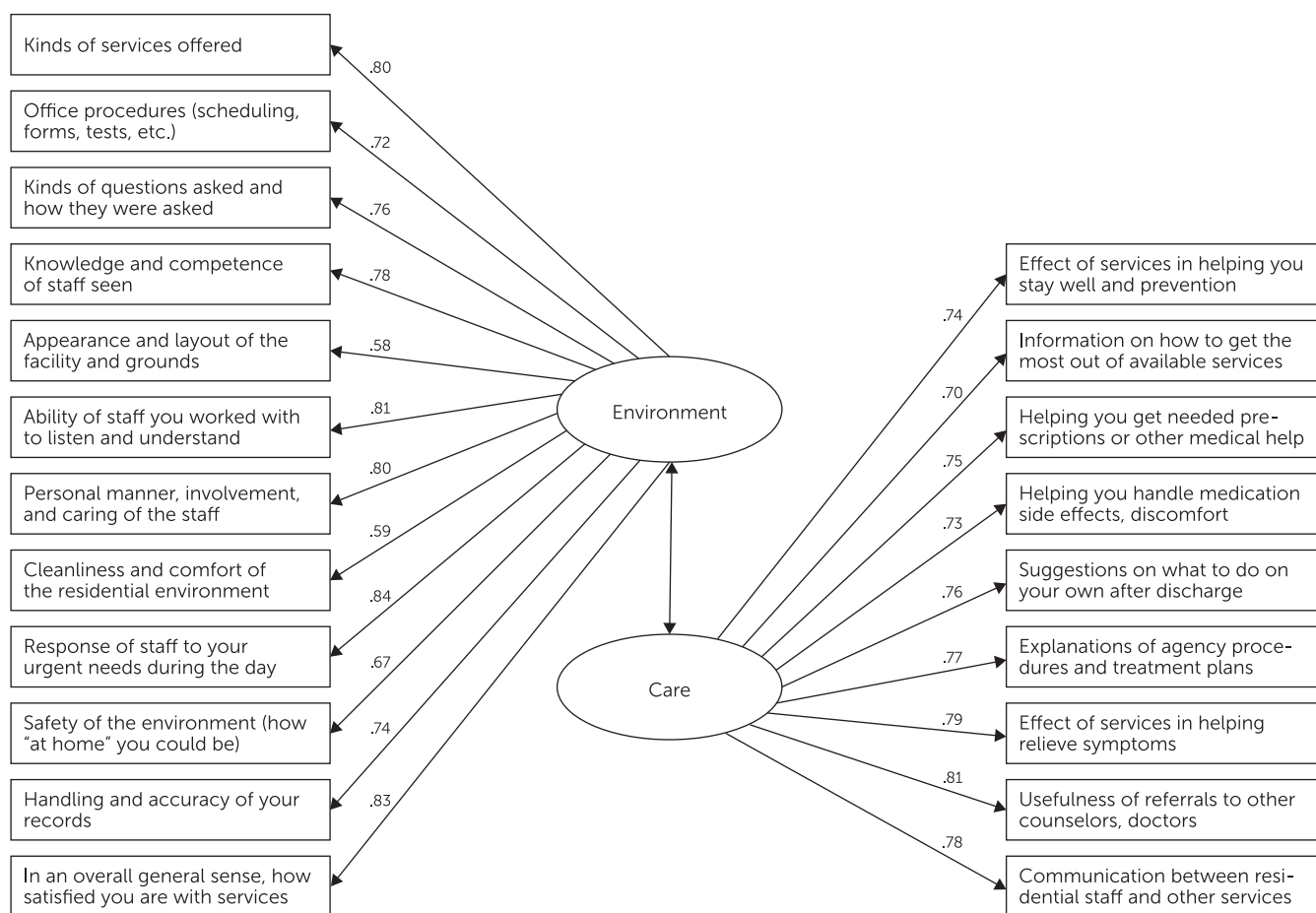


^a Standardized (unit variance) factor loadings are presented (i.e., correlation coefficient values between each of the VOICE items and trust and involvement subscales); larger values indicate higher correlation between the VOICE items and trust and involvement subscales. VOICE, Views on Inpatient Care scale.

the United Kingdom. Similar results with VOICE or SSS-Res are therefore likely to be obtained in other NHS inpatient settings. The more sensitive VOICE measure is likely to reveal more effects on patients' perception of care for new therapeutic activities.

The main strength of this study was that it fully exploited a participatory methodology during PROM development in the context of a trial. Service users were completely involved in instrument development and evaluation throughout the whole research process. Of note, the researchers responsible for data collection and analysis were also mental health service users.

This study was conducted in London boroughs with high levels of deprivation and psychiatric morbidity (27, 28). Our sample included a high proportion of participants from Black and minority racial-ethnic communities, and the sample involved in VOICE development was representative of the local population. Although this involvement was a strength, it may be that different subscales would have been produced by other groups with different sociodemographic backgrounds who would have emphasized different measurement domains of the VOICE instrument.

FIGURE 2. Confirmatory factor analysis of items on the environment and care subscales of SSS-Res for the postintervention sample (N=438 patients)^a

^a Standardized (unit variance) factor loadings are presented (i.e., correlation coefficient values between each one of the SSS-Res items and environment and care subscales); larger values indicate higher correlation between the SSS-Res items and environment and care subscales. SSS-Res, Service Satisfaction Scale: Residential Services Evaluation.

CONCLUSIONS

The VOICE instrument developed and designed by using participatory methods, including service user-led development, was superior to SSS-Res, an instrument previously developed with conventional research methods, in identifying changes in perception of care after staff training on an inpatient ward. The different subscales such as the care scale on SSS-Res were not sensitive to changes in patients' perception of care after the training. Our findings indicate a clear and important impact of involving service users in instrument development and encourage a change in the methods for developing PROMs.

AUTHOR AND ARTICLE INFORMATION

Centre for Implementation Science, Health Services and Population Research Department, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London (Bakolis, Gupta); Department of Biostatistics and Health Informatics (Bakolis) and Department of Psychology (Wykes), Institute of Psychiatry, Psychology and Neuroscience, King's College London, London; South London and Maudsley National

Health Service (NHS) Foundation Trust (Wykes). Send correspondence to Dr. Bakolis (ioannis.bakolis@kcl.ac.uk).

Dr. Bakolis received support from the National Institute for Health and Care Research (NIHR) Biomedical Centre at South London, Maudsley NHS, and from the NIHR Collaboration for Leadership in Applied Health Research and Care South London at King's College Hospital NHS Foundation Trust, King's College London. Drs. Bakolis and Wykes thank the NIHR Biomedical Centre at South London and Maudsley NHS Foundation Trust. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health and Social Care.

The authors report no financial relationships with commercial interests.

Trial registration: ISRCTN 06545047.

Received August 7, 2021; revision received November 17, 2021; accepted January 7, 2022; published online April 27, 2022.

REFERENCES

1. Staniszewska S, Adebajo A, Barber R, et al: Developing the evidence base of patient and public involvement in health and social care research: the case for measuring impact. *Int J Consum Stud* 2011; 35:628–632
2. Staley K: There is no paradox with PPI in research. *J Med Ethics* 2013; 39:186–187

3. Forsythe LP, Ellis LE, Edmundson L, et al: Patient and stakeholder engagement in the PCORI pilot projects: description and lessons learned. *J Gen Intern Med* 2016; 31:13–21
4. Rose D, Evans J, Sweeney A, et al: A model for developing outcome measures from the perspectives of mental health service users. *Int Rev Psychiatry* 2011; 23:41–46
5. Rose D: Participatory research: real or imagined. *Soc Psychiatry Psychiatr Epidemiol* 2018; 53:765–771
6. Katsakou C, Bowers L, Amos T, et al: Coercion and treatment satisfaction among involuntary patients. *Psychiatr Serv* 2010; 61:286–292
7. Evans J, Rose D, Flach C, et al: VOICE: developing a new measure of service users' perceptions of inpatient care, using a participatory methodology. *J Mental Health* 2012; 21:57–71
8. Wykes T, Csapke E, Williams P, et al: Improving patient experiences of mental health inpatient care: a randomised controlled trial. *Psychol Med* 2018; 48:488–497
9. Penn DL, Roberts DL, Combs D, et al: Best practices: the development of the social cognition and interaction training program for schizophrenia spectrum disorders. *Psychiatr Serv* 2007; 58:449–451
10. Reeder C, Pile V, Crawford P, et al: The feasibility and acceptability to service users of CIRCuiTS, a computerized cognitive remediation therapy programme for schizophrenia. *Behav Cogn Psychother* 2016; 44:288–305
11. Kavanagh K, Duncan-Macconnell D, Greenwood K, et al: Educating acute inpatients about their medication: is it worth it? An exploratory study of group education for patients on a psychiatric intensive care unit. *J Mental Health* 2003; 12:71–80
12. Ruddle A, Mason O, Wykes T: A review of hearing voices groups: evidence and mechanisms of change. *Clin Psychol Rev* 2011; 31:757–766
13. Linehan MM, Heard HL, Armstrong HE: Naturalistic follow-up of a behavioral treatment for chronically parasuicidal borderline patients. *Arch Gen Psychiatry* 1993; 50:971–974
14. Grey S: Problem solving groups for psychiatric inpatients: a practical guide; in *Psychological Groupwork With Acute Psychiatric Inpatients*. Edited by Radcliffe J, Hajek K, Caron J, et al. London, Whiting & Birch, 2010
15. Knight MTD, Wykes T, Hayward P: Group treatment of perceived stigma and self-esteem in schizophrenia: a waiting list trial of efficacy. *Behav Cogn Psychother* 2006; 34:305–318
16. Greenfield T, Attkisson C: The UCSF client satisfaction scales: II. The Service Satisfaction Scale-30; in *The Use of Psychological Testing for Treatment Planning and Outcomes Assessment: Vol 3: Instruments for Adults*, 3rd ed. Edited by Maruish ME. Mahwah, NJ, Erlbaum, 2004
17. Muthen B, du Toit S, Spisic D: Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. *Psychometrika* 1997; 75:40–45
18. Bentler PM: Comparative fit indexes in structural models. *Psychol Bull* 1990; 107:238–246
19. Tucker LR, Lewis C: A reliability coefficient for maximum likelihood factor analysis. *Psychometrika* 1973; 38:1–10
20. Steiger JH: Notes on the Steiger–Lind (1980) Handout. *Struct Equ Modeling* 2016; 23:777–781
21. Hu L-t, Bentler PM: Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Modeling* 1999; 6:1–55
22. Cronbach LJ: Coefficient alpha and the internal structure of tests. *Psychometrika* 1951; 16:297–334
23. Hemming K, Girling A, Martin J, et al: Stepped wedge cluster randomized trials are efficient and provide a method of evaluation without which some interventions would not be evaluated. *J Clin Epidemiol* 2013; 66:1058–1059
24. Barker D, McElduff P, D'Este C, et al: Stepped wedge cluster randomised trials: a review of the statistical methodology used and available. *BMC Med Res Methodol* 2016; 16:69
25. Hussey MA, Hughes JP: Design and analysis of stepped wedge cluster randomized trials. *Contemp Clin Trials* 2007; 28:182–191
26. Bird V, Miglietta E, Giacco D, et al: Factors associated with satisfaction of inpatient psychiatric care: a cross country comparison. *Psychol Med* 2020; 50:284–292
27. Kirkbride JB, Morgan C, Fearon P, et al: Neighbourhood-level effects on psychoses: re-examining the role of context. *Psychol Med* 2007; 37:1413–1425
28. Morgan C, Dazzan P, Morgan K, et al: First episode psychosis and ethnicity: initial findings from the AESOP study. *World Psychiatry* 2006; 5:40–46