Patient Preference and Adherence

Dovepress
open access to scientific and medical research

Open Access Full Text Article

STUDY PROTOCOL

The Patient-Reported Experience Measure for Improving qUality of care in Mental health (PREMIUM) project in France: study protocol for the development and implementation strategy

This article was published in the following Dove Medical Press journal:
Patient Preference and Adherence

Sara Fernandes,[1] Guillaume Fond,[1] Xavier Zendjidjian,[1] Pierre Michel,[1] Karine Baumstarck,[1] Christophe Lancon,[1] Fabrice Berna,[2] Franck Schurhoff,[2] Bruno Aouizerate,[2] Chantal Henry,[2] Bruno Etain,[2] Ludovic Samalin,[2] Marion Leboyer,[2] Pierre-Michel Llorca,[2] Magali Coldefy,[3] Pascal Auquier,[1] Laurent Boyer[1]

On behalf of the French PREMIUM Group

[1]Aix-Marseille University, School of Medicine, CEReSS – Health Service Research and Quality of Life Center – EA 3279 Research Unit, Marseille, France; [2]FondaMental Foundation, Créteil, France; [3]Institute for Research and Information in Health Economics (IRDES), Paris, France

Correspondence: Sara Fernandes
Aix-Marseille Université, School of Medicine – La Timone Medical Campus, EA 3279: CEReSS – Health Service Research and Quality of Life Center, 27 Boulevard Jean Moulin, 13005 Marseille, France
Tel +33 66 018 5077
Email sfernandes.sara@gmail.com

Background: Measuring the quality and performance of health care is a major challenge in improving the efficiency of a health system. Patient experience is one important measure of the quality of health care, and the use of patient-reported experience measures (PREMs) is recommended. The aims of this project are 1) to develop item banks of PREMs that assess the quality of health care for adult patients with psychiatric disorders (schizophrenia, bipolar disorder, and depression) and to validate computerized adaptive testing (CAT) to support the routine use of PREMs; and 2) to analyze the implementation and acceptability of the CAT among patients, professionals, and health authorities.

Methods: This multicenter and cross-sectional study is based on a mixed method approach, integrating qualitative and quantitative methodologies in two main phases: 1) item bank and CAT development based on a standardized procedure, including conceptual work and definition of the domain mapping, item selection, calibration of the item bank and CAT simulations to elaborate the administration algorithm, and CAT validation; and 2) a qualitative study exploring the implementation and acceptability of the CAT among patients, professionals, and health authorities.

Discussion: The development of a set of PREMs on quality of care in mental health that overcomes the limitations of previous works (ie, allowing national comparisons regardless of the characteristics of patients and care and based on modern testing using item banks and CAT) could help health care professionals and health system policymakers to identify strategies to improve the quality and efficiency of mental health care.

Trial registration: NCT02491866.

Keywords: patient-reported experience measures, quality of care, item bank, computerized adaptive testing, psychiatry

Introduction

Mental disorders affect on average one in five adults,[1] are leading causes of disability worldwide,[2] and are associated with premature mortality and excess costs.[3] Poor quality has been reported in the diagnosis, treatment, and follow-up of patients with mental disorders such as schizophrenia, bipolar disorder, or depression.[4] These mental disorders are often unrecognized or misdiagnosed, leading first to prolonged duration of untreated psychosis and depression[5–7] and subsequently to poor outcomes in treatment response, symptoms, and quality of life.[8] Under-use of guidelines and inadequate or suboptimal treatments,[9] health care variation among geographical regions,[10] and poor

submit your manuscript | www.dovepress.com
Dovepress
http://dx.doi.org/10.2147/PPA.S172100

adherence to treatment by patients remain major challenges for mental health care.[11–13] There is thus a need to measure the quality and performance of mental health care in France[14,15] as in other Western countries,[16,17] in order to propose strategies to improve the quality and efficiency of mental health care.[18]

Patient experience is considered to be one important measure of health care quality.[19–21] The use of patient-reported experience measures (PREMs) is recommended by the Organization for Economic Co-operation and Development.[22,23] Relationships among patient experience, the process of care, and health outcomes are well recognized. In mental health, patient experience of care is predictive of future behaviors, including intent to return for care, promptness in seeking help for further episodes, adherence to treatment, and quality of life.[24–28] Many PREM questionnaires in mental health have been developed over the past decade,[29–33] but they address a condition- or care-specific group of patients (eg, in- or outpatient,[34,35] people with specific illnesses,[36,37] one type of psychiatric care).[38] This specificity makes general assessments and comparisons at a national or international level difficult. In addition, most available PREMs are paper-based, making it challenging for professionals to obtain quality of care scores efficiently in real time. The questionnaires are frequently too lengthy and fixed in content (ie, asking the same questions to all patients regardless of their health characteristics), leading to a high survey burden for patients and to substantial problems with missing data.[39] As a consequence, PREMs are not routinely collected in France, and assessment of the quality of mental health care remains mainly based on statistics from national administrative databases[10,40–45] and indicators of patient record keeping.[46] In addition, novel approaches and reimbursement systems are currently being tested[14,15,47,48] and may have profound effects on the mental healthcare system in France. Their effects, including on patients' perceptions and needs, need to be monitored accurately and scientifically. In this context, the Patient-Reported Experience Measure for Improving qUality of care in Mental health (PREMIUM) Group received funding from the French National Health System's research programme on the performance of health care system (PREPS) n°13–0091 QDSPsyCAT to develop a set of PREMs on the quality of care in mental health. In particular, the project seeks to overcome limitations such as the possibility of making national comparisons regardless of the characteristics of patients or the care received. The PREMIUM Group seeks to develop the new PREMs based on modern testing methods, including item banks and computerized adaptive testing (CAT) that are already used for health outcomes and patient-reported outcome measures (PROMs) in psychiatry.[49–52] First, PREMs allow patients to provide direct feedback on their care to drive improvement in services and are complementary to PROMs which capture a person's perception of their health.[53] The combined use of patient reports advances the patient-centered healthcare approach.[54,55] Second, CAT is based on the item response theory (IRT) and allows the administration of a customized subset of items taking into account the candidate's level of ability for the latent trait being studied. Thus, only the most suitable items to assess the quality of care perceived by the respondent will be administered. As such, it provides more accurate score estimates and represents a lower burden than standard fixed format questionnaires.[56]

The aim of this project is therefore 1) to develop item banks of PREMs on quality of care in mental health applicable to adult patients with mental health disorders (ie, schizophrenia, bipolar disorder, and depression) and to validate CAT in order to support the routine use of PREMs and 2) to analyze the implementation and acceptability of this tool among patients, professionals, and health authorities.

Methodology

Study design

This multicenter and cross-sectional study is based on a mixed method approach associating qualitative and quantitative methodologies. It follows two phases:

1. Item banking and CAT development based on a standardized procedure:[57–60] conceptual work and definition of the domain mapping, item selection, calibration of the item bank and CAT simulations for the elaboration of the administration algorithm, and CAT validation.
2. Qualitative study exploring implementation and acceptability of this tool among patients, professionals, and health authorities.

Figure 1 shows the study flow chart.

The PREMIUM group, project organization, and study settings

The PREMIUM group aims to promote and facilitate opportunities to develop and use PREMS in mental health research and care in France. This group is a multidisciplinary and interprofessional team composed of representatives from patients' associations, public health experts, psychiatrists, psychologists, economists, biostatisticians, mathematicians, and programmers from different research teams (Center on Health Service Research CEReSS, Aix-Marseille University, Marseille, France; I2 M UMR 7373 – Mathematics Institute

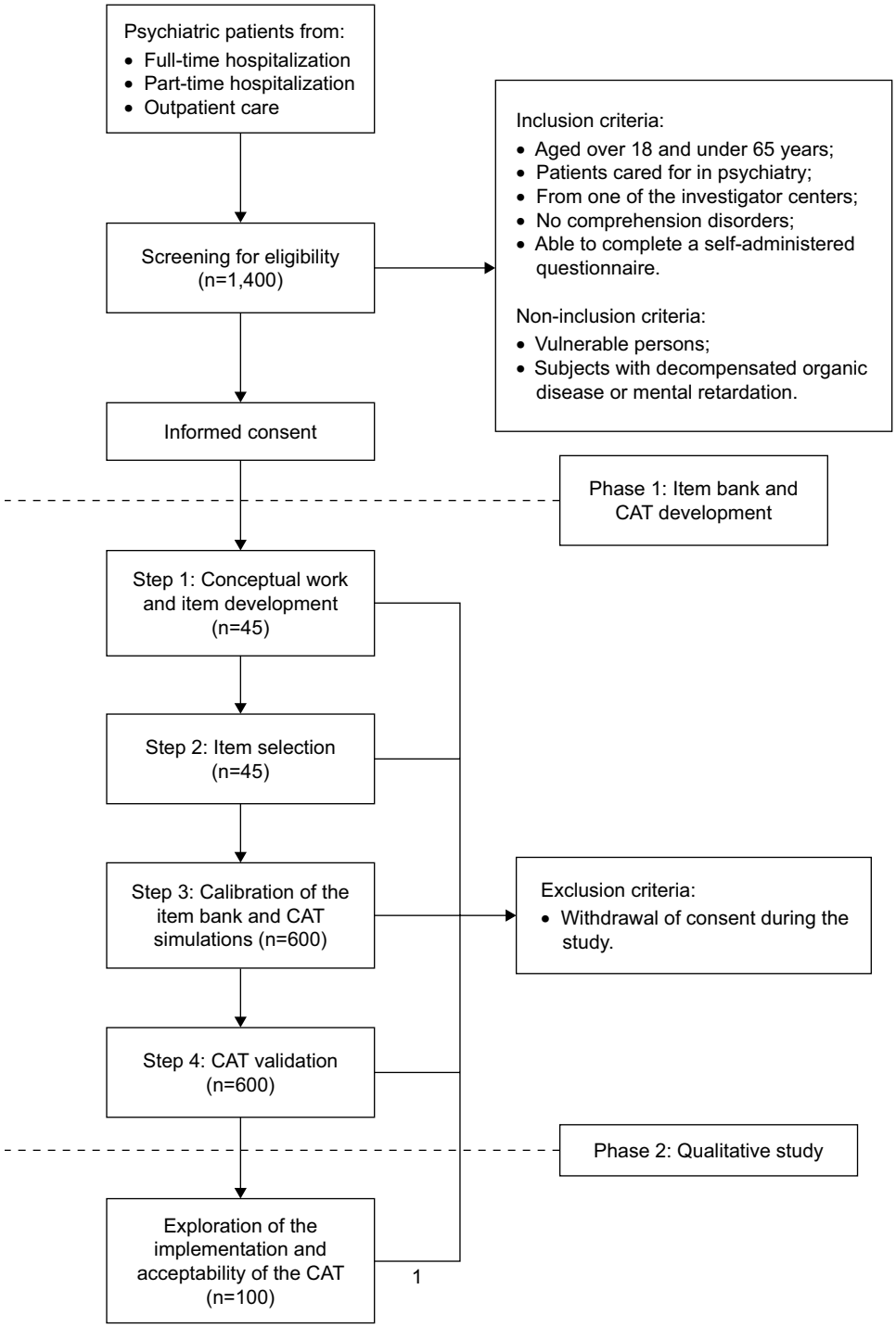

Figure 1 Study flow chart.
Abbreviation: CAT, computerized adaptive testing.

of Marseille; EA 7280 – University of Auvergne; INSERM U955; Fondation FondaMental; Institut de recherche et documentation en économie de la santé [IRDES]).

The PREMIUM organizational structure is composed of a steering committee and five executive committees. The steering committee governs the project and is in charge of validating the different steps of the project including the conceptualization and metrological validation of the PREMs. The five executive committees are as follows: item bank development team, recruitment team, data management and psychometric analysis team, software development team, and communication team.

The following clinical sites (including full-time hospitalization, part-time hospitalization, and ambulatory care settings) throughout the French territory will be involved in the recruitment of participants: Assistance Publique – Hôpitaux de Marseille, Assistance Publique – Hôpitaux de Paris, Centre Hospitalier de Toulon, Centre Hospitalo-Universitaire Clermont-Ferrand and the French network of expert centers (Fondation Fondamental) for schizophrenia (10 centers), bipolar disorder (10 centers), and depression (13 centers).[14,15,61]

Patient screening will be performed by the investigators of the centers included in this study to ensure that patients who meet the inclusion criteria are correctly identified.

The protocol and purpose of the study will be explained orally and in written form to each participant in order to obtain their informed consent. Patients will be informed that their participation is voluntary and that they can withdraw from the study at any time. Participants will also be assured of the anonymity of their answers.

Inclusion and exclusion criteria

Details of the inclusion and exclusion criteria for the two phases are provided in Table 1.

Study procedure

Phase 1: item banking and CAT development:

This first phase involves four steps:

1. Conceptual work and definition of the domain mapping: face-to-face semi-structured interviews will be conducted with patients (see inclusion and exclusion criteria in Table 1) to define a domain map describing mental health quality of care based on the patient point of view.
2. Item selection: this step begins with a systematic review to identify existing items in currently available PREMs in mental health. A standardized item library will collect the following characteristics: author, date of validation; country of origin, language; title of the PREM; context of use (eg, condition- or care-specific focus); the dimensions or domains of the questionnaire; the items; instructions associated with answering items; response options; time frame; response rate; and instrument availability. After identification and item collection, all items will be translated into the French language following international guidelines.[62,63] Then, the PREMIUM Group experts will: 1) select the most understandable and representative items (ie, remove redundant, ambiguous, and difficult items); 2) review and revise each item to provide consistency in style (for items, response option, and time frame); and 3) classify the items according to the domains identified during the previous step. Finally, face-to-face semi-structured interviews will be conducted with patients on all the selected items to elicit feedback on language, understandability, unambiguity, the relevance of each item, response option and time frame, and any omissions of important information on quality of care. All comments will be taken into account in the correction process.

Table 1 Selection criteria

Inclusion criteria	
Phase 1 (item bank and CAT development)	Men or women, aged over 18 and under 65 years
	Subjects cared for in one of the investigator centers of this study
	Cared for in psychiatry; diagnosis of schizophrenia, bipolar disorder or depression according to DSM V, regardless of current or previous therapies, duration, and severity of illness
	Subjects who have no comprehension disorders and are able to read and write, agree to participate in the study, and have signed informed consent
Phase 2 (qualitative study)	Any patient or health system stakeholder from an investigator health care establishment or an associated supervisory organization: professionals involved in the care of patients (doctors, nurses, and psychologists), department heads, health executives, members of the hospital administration, and public institutions
Exclusion criteria	
	Aged less than 18 or greater than 65 years
	Subject not cared for in psychiatry or untreated in one of the investigative centers, without a diagnosis of schizophrenia, bipolar disorder, or depression according to DSM V
	Subject unlikely to participate in an interview or complete a self-administered questionnaire
	Subject refusing to participate in the study or to sign informed consent
	Vulnerable persons (pregnant women, parturient or nursing mothers, persons deprived of liberty, persons admitted to a health or social institution for other purposes than research, and adults who are subject to a legal protection measure)
	Subjects with decompensated organic disease or mental retardation
	Subject withdrawing his/her consent before the end of the study

Abbreviations: CAT, computerized adaptive testing; DSM V, *Diagnostic and Statistical Manual of Mental Disorders*, fifth edition.

Items that are ambiguous or misunderstood will be removed or reworded; new items could be added if important missing subjects are highlighted by patients. Each step, from the literature research to the final list of items, will be performed by two independent reviewers, and a third reviewer will be involved in case of disagreement.

3. Calibration of the item bank and CAT simulations for the elaboration of the administration algorithm: item bank calibration is the prerequisite for developing CATs.[64,65] The list of items will be tested on a large and heterogeneous sample of patients with mental health disorders (ie, different diagnoses, care settings, and cities) to choose the most appropriate IRT model fitting into the data and check for skewness, unidimensionality, local independence, differential item functioning (DIF), and item fit. According to the findings, some items can be discarded (eg, in case of violation of the assumption of monotonicity, local independence, significant DIF, or poor fit indices). A real-data simulation approach (ie, complete response patterns to all the selected items) will be used to simulate the conditions of the CAT assessment. We will use the responses contained in the item banks to simulate the adaptive administration of items. The principle of CAT simulations is presented in Figure 2. At the end of this analysis, the best item administration algorithm will be defined using different scenarios of computerized adaptive tests and simulated data.
4. CAT validation will be performed on a large and heterogeneous new sample of patients who will fulfill the CATs. Complementary data will be collected to explore the clinical relevance of the CATs, particularly by testing the link between the CATs and potentially related concepts such as satisfaction, therapeutic alliance, severity of symptoms, and quality of life. Acceptability indicators will be computed to test the relevance of the measure (percentage of missing data and average completion time).

Phase 2: a qualitative study exploring the implementation and acceptability of the CAT will be conducted using face-to-face semi-structured interviews with patients, professionals, and health authorities.

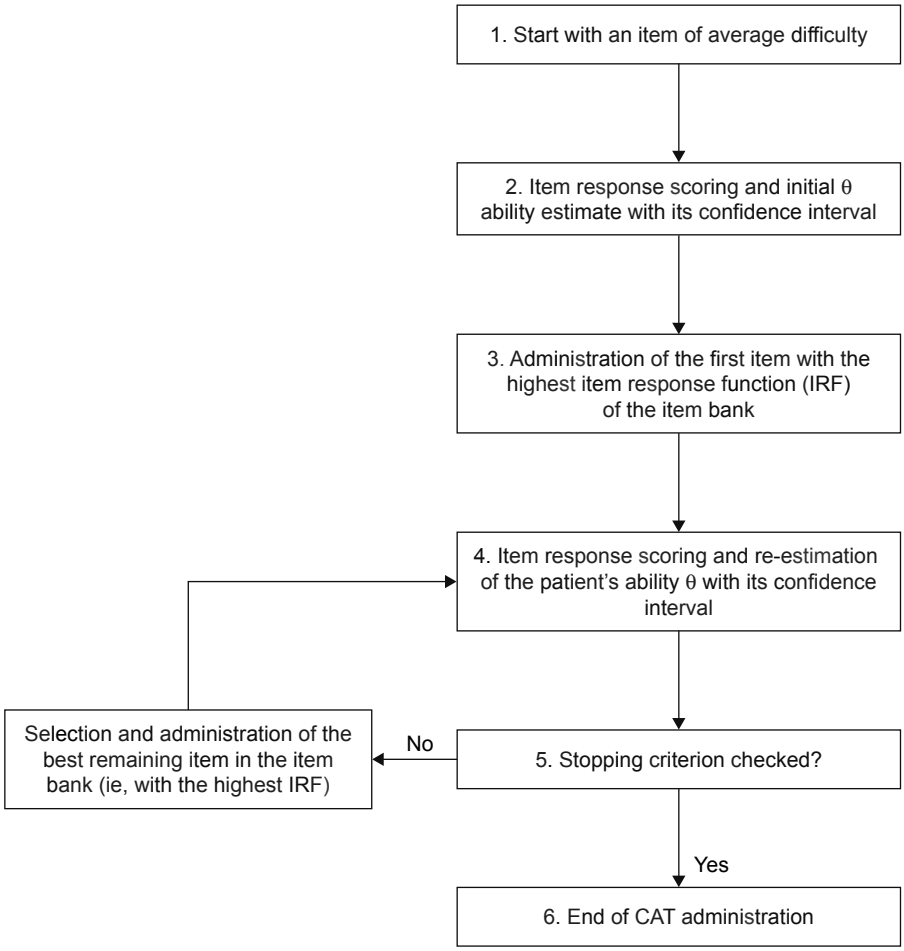

Figure 2 CAT algorithm.
Abbreviation: CAT, computerized adaptive testing.

Table 2 Collected data to assess the quality of mental health care

Measure	Instruments	Number of items and format	Number of dimensions	Description	Steps of Phase 1*
Sociodemographic characteristics	NA	NA	NA	Gender, age, educational level, marital status, work situation	1, 2, 3, and 4
Clinical data	NA	NA	NA	Disease duration, psychiatric diagnosis	1, 2, 3, and 4
Severity	Clinical Global Impression-Severity (CGI-S)[123,124]	1 item, 7-point Likert	1	Clinician's assessment of current severity of the patient's symptoms	4
General functioning	General Assessment Functioning (GAF)[125]	Numeric scale (ranged from 1 to 100)	1	Clinician's assessment of the individual's overall level of functioning (social, occupational, and psychological)	4
Satisfaction	Client Satisfaction Questionnaire[126]	8 items, 4-point Likert	1	Client's assessment of satisfaction with mental health services	4
Therapeutic alliance	4-point ordinal Alliance Self-report[127]	11 items, 4-point Likert	2	Assessment of the quality of the therapeutic relationship	4
Medication adherence	Medication Adherence Rating Scale[128]	10 items, Yes/No	3	Assessment of the patient's medication adherence	4
Quality of life	The Short-Form (36) Health Survey[129]	36 items	8	Assessment of a person's self-report quality of life	4

Notes: *Phase 1 includes 4 steps as described in the "Study procedure" section. All 4 steps were considered for some measures, whilst step 4 only was considered for other measures.
Abbreviation: NA, not applicable.

Data collection

The collected data are presented in Table 2.

Sample size

The sample size calculation is presented in Table 3.

Statistical considerations

Descriptive analysis

The distribution of item response categories will be described using the mean and standard deviation. Floor or ceiling effects will also be studied. Some items may be excluded if: 1) high (>70%) missing value rates; 2) extreme skewness (>95% responses in one category); and 3) interitem correlation coefficients evaluated by Spearman's nonparametric correlation higher than 0.70, which indicates some redundancy between these items.[58,66]

Evaluation of the assumptions of IRT model

Dimensionality will be evaluated using exploratory and confirmatory factor analysis (CFA) methods for categorical data. Analyses will be conducted assuming a single latent dimension for each item bank domain.

The factorability of the dataset will be evaluated by the Kaiser–Meyer–Olkin test and Bartlett's sphericity test.[67]

Table 3 Sample size

	Sample size	Justification
Phase 1		
Step 1	45 face-to-face semi-structured interviews	In the absence of basic guidelines, the concept of saturation[130] is considered the guiding principle of qualitative research to determine sample size.[131] We chose to include 45 subjects in accordance with the recommendations of Ritchie et al[132] that an optimal sample should contain less than 50 subjects and that beyond 60 inclusions data collection does not provide additional information necessary for understanding the purpose of the study.
Step 2	A minimum of 45 face-to-face directive interviews	
Step 3	600 patients	There is no consensus on sample size requirements for estimating the parameters of an IRT model.[133,134] However, some guidelines have been formulated. In accordance with those formulated by Tsutakawa and Johnson,[135] a sample higher than 500 individuals is suitable to obtain sufficiently accurate estimates when a multi-parameter model is used.[136]
Step 4	600 patients	There is a lack of formal guidelines on how to calculate a priori sample size for psychometric validation studies.[137] The minimum size depends on the methods considered. To obtain a heterogeneous sample covering the spectrum of possible responses, we chose to include 600 topics as in the previous step (1.3).
Phase 2	A minimum of 45 patients and 45 health stakeholders	The number of subjects to be included in this second phase was determined according to the recommendations based on the saturation concept used in steps 1.1 and 1.2.

Thereafter, a principal component analysis will be performed, followed by a CFA to validate the structure of the model being studied.[68] A one-factor CFA will be compared to a bifactor model to explain potential deviations from the unidimensionality assumption.[69,70] Several criteria can be used to determine the number of factors to extract: the cumulative percent of variance explained, the Kaiser-Guttman's rule (eigenvalues ≥ 1),[71] the scree test (looking for an "elbow" in the curve),[72] and parallel analysis.[73] Items with factor loading below 0.40 (or in some cases below 0.30 to ensure content validity)[74] will be discarded.[75] Model fit will be evaluated with commonly used model fit indices: the root mean square error of the approximation with values below 0.05 indicating a good fit, values between 0.05 and 0.08 reflecting an adequate fit, and values greater than 0.08 meaning a marginal fit. The Tucker–Lewis index (TLI) or the comparative fit index (CFI) with values ≥ 0.90 suggests reasonable fit, while TLI/CFI values ≥ 0.95 reflect a good fit of the model to the data.[76–79]

Internal consistency will be investigated by Cronbach's alpha coefficient with $\alpha > 0.7066$ considered as acceptable.

Local independence is characterized by the absence of a significant relationship between item responses when the ability's level is controlled.[80] This prerequisite will be explored by analyzing the matrix of residual correlations with strong correlations suggesting the existence of a local dependence. If a pair of items has a residual correlation ≥ 0.20 or ≥ 0.25,[81,82] the item with the highest cumulative residual correlation with the remaining items will be eliminated.

Monotonicity postulates that the probability of "success" (or endorsement) of an item increases with a person's ability level. This relationship is modeled by a monotonous, non-decreasing function and can be visually verified using the item characteristic curves. Analysis of the item characteristic curves will also verify that each response category of an item has a maximum probability of being selected on a specific range of the latent trait scale. If two options for responding to an item are not sufficiently discriminatory, they will be collapsed, and the model will be readjusted accordingly.[81,83,84]

Calibration and fitting of an IRT model to the data

Item parameters will be estimated from an IRT model fitted to polytomous data: the generalized partial credit model (GPCM) by Muraki[85] and the maximum marginal likelihood estimation procedure.[86] In the GPCM, the items do not necessarily have the same number of response categories and are not ordered. Each item is characterized by a slope parameter (ie, item discrimination between individuals with different ability levels) and by a set of threshold parameters (ie, item difficulty). This two-parameter model has the advantage of allowing slope parameters to vary across items, and threshold parameters between response categories give an indication on the response options' locations along the latent construction continuum. This model is widely used in health research[64,82,87,88] and allows for a more accurate description of data than the one-parameter partial credit model from which it derives and in which all items are equally discriminating.[88] To evaluate the relevance of the model chosen, the data will also be calibrated by the graded response model (GRM),[89] which is known to offer a similar fit to the data.[89–91]

The goodness of fit of each item to the model will be examined through the Infit Mean Square (Infit MnSq) statistic, which evaluates the correspondence between the expected and observed response models. Infit MnSq is more affected by unexpected responses to items close to the person's ability level. The range of 0.6–1.4 is considered acceptable, with a better fit to the theoretical model when the Infit MnSq is close to 1.[92]

Evaluate DIF

DIF is a systematic error in the functioning of an item that occurs when there is an interaction between belonging to a subgroup (such as sex or age) for individuals with the same level of ability and the response to a particular item.[93,94] Failure to consider a DIF can interfere with the measurement validity. DIF analysis will be performed using an IRT-based iterative ordinal logistic regression method. To do this, a GRM-type IRT model will be used because of its inherent connection to ordinal logistic regression.[95]

Evaluation of the DIF magnitude will be done according to Zumbo's classification using pseudo-R^2 measures (ΔR^2 with a negligible DIF if $\Delta R^2 < 0.13$, moderate if $0.13 < \Delta R^2 < 0.26$, and large if $\Delta R^2 > 0.26$).[96] Items with a large DIF will be discarded, while those with a moderate DIF will be discussed.

Elaboration of items administration algorithm

CAT simulations will be performed using a "real data simulation" method from the full sample of participants of the previous step (ie, calibration of the item bank) by cross-validation method. Several CAT administration scenarios will be created and compared in order to select the most powerful algorithm based on predetermined stopping criteria. The goal will be to find an optimal balance between the accuracy of the scores and the respondents' burden.

The algorithm will start by estimating an initial average score θ for each individual, according to which the algorithm will select and administer the item with the highest information function in the bank. Score θ and its CI will be re-estimated iteratively on the basis of the response to the previous item. We will use the expected a posteriori method for scoring.[97] The precision of the CAT (ie, the accuracy of the IRT-based score estimation) will be assessed against scores based on the full responses of the item bank using root mean square errors, for which a value of 0.3 or less means excellent measurement precision.[98] Empirical reliability also depends on the standard error of measurement (SEM). The lower the SEM, the higher the CAT reliability.[99] To achieve satisfactory reliability (≥0.70), the SEM must be less than 0.55.[100]

Figure 2 shows the CAT algorithm adapted from Wainer et al.[101]

An alternative to IRT-based CAT using machine learning and decision trees will also be tested in accordance with recent work on this issue.[102]

Validation of the CAT

Divergent validity will be tested by comparing the mean scores by dimension between groups of patients for whom assumptions of difference can be made based on their sociodemographic (ie, age, gender, academic level, marital status, and work situation) and clinical (ie, disease duration, Clinical Global Impression-Severity and General Assessment Functioning scores) characteristics using Student's *t*-tests, ANOVAs, Pearson correlations, and post hoc analyses. One general assumption is that the more a patient's clinical condition deteriorates, the lower his CAT score should be.

Convergent validity will be determined by investigating the correlation between the mean scores per item bank domain and those of instruments supposed to indirectly measure the concept of quality of care.[103–106] Pearson's correlation coefficients will be calculated, and we will assume a positive correlation between the scores of scales exploring similar concepts.

Test–retest stability will be evaluated using intraclass correlation coefficients (ICCs) between the responses made by the same individuals at 15-day intervals to limit memory bias. ICC values ≤0.40 will be considered insufficient, values between 0.41 and 0.60 will be considered moderate, values from 0.61 to 0.80 will be considered good, and values >0.81 will be considered excellent.[107]

Descriptive statistics will explore the acceptability of the instrument. Group comparisons using Student's *t*-tests, ANOVAs, and post hoc (Tukey-type) tests will be performed to determine if there are significant differences in the distribution of missing data based on sample characteristics.

Qualitative analysis

The discourses of different actors in the health system (patients, professionals, and health authorities) will be analyzed using two complementary approaches. As a first step, the transcripts of interviews will be subjected to a thematic content analysis. Two researchers will independently read and code the interviews to identify aspects deemed important for quality care from the patient's perspective. The interviews will also be subjected to computerized text analysis. Researchers will compare and discuss the results to reach consensus on findings.

Registry and ethical approval

The trial registration is NCT02491866. At the time of manuscript submission, the status of the trial is recruiting.

The study is being carried out in accordance with ethical principles for medical research involving humans.[108] The assessment protocol was approved by the relevant ethical review board (CPP-Sud Méditerranée V, November 12, 2014, n°2014-A01152-45). All data are collected anonymously. As this study includes data coming from regular care assessments, a nonopposition form was signed by all participants.

Discussion

To our knowledge, the PREMIUM study is the first study to propose the development of a common measurement system for assessing patient-reported experience of mental health care. In recent years, various common measurement systems for health care performance assessment have been developed in European countries,[109] such as the Quality Indicator for Rehabilitative Care – QuIRC,[110,111] the Measure of Best Practice for People with Long Term Mental Illness in Institutional Care – DEMOBinC,[112] Quality Monitoring Programmes for Mental HealthCare (QMP-MHC),[113] and the Description and Evaluation of Services for Long Term Care in Europe (DESDE-LTC) from the recent research on Financing Systems' Effect on the Quality of Mental Health Care in Europe.[114] However, these measurements mainly focus on availability, diversity, and capacity of mental health care resources; they do not include "what matters to patients".[21] Other initiatives have been proposed to consider patients' views, such as the patient-reported outcomes measurement information system[115] and the International Consortium for Health Outcomes Measurement.[116] However,

these measure patients' outcomes of health (ie, PROMs). This project is therefore complementary to these other initiatives. It recognizes the importance of integrating patients' experiences of their care into mental health care assessment and health research services.

This work is expected to be of great interest in France, where significant regional disparities in the mental health care system have been reported, without significant changes, over recent decades.[117] According to experts, a reallocation of resources between psychiatric institutions is urgently needed to guarantee the quality of and equity in access to mental health care in France.[48] Adopting a common standard and metric will enable us to directly compare patients' views of current delivery and settings of mental health care in France. Standardized PREMs could thus become a key component of a national reflection on the mental health care system in France. This work may also be exported to foreign countries. As this project will propose PREMs based on a comprehensive systematic review of all existing items in available PREMs and patients' perspectives (extracted from interviews on the domain mapping and final selected items), it will provide internationally replicable measures that will allow direct comparisons of mental health care systems. Generating a common set of standardized PREMs that can be utilized widely by the international community has great potential to contribute to developing health service research in mental health and ultimately improving health care worldwide.

This study faces several challenges. Even with a large overall sample size for this multicenter study, the sample may not be representative of the population with schizophrenia, bipolar disorder, and depression. Because the study is taking place in large centers, its findings may not be generalizable to patients in smaller centers where care, life conditions, and needs may be different. However, our study includes centers in cities across France, thus taking into consideration at least some potential health care, socioeconomic, and cultural differences.

Each mental illness may be associated with specific needs. Presently, the items are being selected to enable comparisons among schizophrenia, bipolar disorder, and depression. The development of item banks for these three initial disorders occurs in the context of severe and persistent mental illnesses that share similarities. Inclusion is limited to these three main psychiatric disorders due to the need for a homogeneous population and challenges in managing these illnesses in the context of low quality reported in the diagnostic, treatment, and follow-up of these patients.[4] This work will be extended to other psychiatric illnesses in the future (including other Axis I and II diagnoses). People aged 65 years and over were also excluded from this study because they have specific issues that differ from those of working-age adults. This study may be replicated in the future for older adults.

Another possible limitation would be the item bank's ability to comprehensively cover the concept to be measured. Given frequent cognitive impairment in the target population, relatively short item banks – with a maximum of about 30 items per domain – would be preferable.[118] However, the "exhaustiveness" of the item bank then becomes questionable. To address this issue, patient interviews will include seeking feedback on any important aspects of quality of care that have not been covered, and the contents of the item bank will be modified based on their comments. Thus, the item bank should be sufficiently extensive to comprehensively cover the concept of care quality in mental health.

A trend toward high levels of positive ratings in surveys designed to measure patient experience is another well-described issue.[119–121] This trend is particularly true for people suffering from severe mental illness[122] due to various factors (social desirability bias, low involvement of respondents in the item generation process, fear of an impact on care, and others). To limit this, the CAT will be developed using a mixed methodology and including patients at all stages of instrument development. Moreover, all participating patients are informed that the questionnaires are anonymized, will not be returned to their treating psychiatrist, and therefore will not impact their care. The expert center network appears as the best structure to address this issue in France,[14] as the questionnaire will be administered in expert centers and not in ambulatory health care settings, to preserve the independence of the evaluation. However, despite these precautions, this problem may persist and should be taken into account in interpreting the results.

Finally, the use of CATs has many well-known advantages. The routine use of patient reports requires the implementation of a measurement system that is accurate and robust, yet acceptable and suitable for clinical use. CAT meets many of these guidelines. The substantial time savings achieved by administering a CAT are more appropriate to the clinical context. Patients will be more likely to respond to items that take into account their characteristics. CAT improves accuracy and precision of scores, which make PREMs more relevant for professionals to improve quality of care and services. However, computerization can also raise problems in routine use. The implementation and use of a CAT assumes the existence of dedicated computers, which

is not necessarily the case in health care settings. Moreover, reliance on CATs implies basic computer literacy among all respondents, introducing a risk of placing particular respondents (such as people with a lower education level or significant cognitive impairments) at a disadvantage. These elements could create barriers to effective use of CATs. To identify potential barriers to widespread use of CATs, an acceptability study will be conducted to propose interventions to optimize their use.

Conclusion

The PREMIUM project is the first to develop an innovative tool to assess quality of care in mental health services. Its work could be extrapolated to other countries in the future, enabling comparisons across countries. Three psychiatric disorders have been selected as the focus in the development of questionnaires; these questionnaires will likely be replicated for other mental illnesses going forward. These PREMs will provide information that could help support public decision-making and improve the transparency and organization of health care, as well as professional practice. They will also support the provision of appropriate care for people with mental disorders in the light of scientific knowledge and with respect for patients' expectations and needs.

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Acknowledgments

This work is supported by an institutional grant from the The French National Program on the Performance of the Healthcare System (PREPS, financed by Direction Générale de l'Offre de Soins, 14, avenue Duquesne, 75350 Paris, France). The sponsor has no role in study design; collection, analysis and interpretation of data; report writing; or the decision to submit the article for publication.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Steel Z, Marnane C, Iranpour C, et al. The global prevalence of common mental disorders: a systematic review and meta-analysis 1980–2013. *Int J Epidemiol*. 2014;43(2):476–493.
2. Whiteford HA, Degenhardt L, Rehm J, et al. Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. *Lancet*. 2013;382(9904):1575–1586.
3. Charlson FJ, Baxter AJ, Dua T, Degenhardt L, Whiteford HA, Vos T. Excess mortality from mental, neurological and substance use disorders in the Global Burden of Disease Study 2010. *Epidemiol Psychiatr Sci*. 2015;24(2):121–140.
4. Institute of Medicine. *Improving the Quality of Health Care for Mental and Substance-Use Conditions*. Washington, DC: National Academies Press; 2006.
5. Baca-Garcia E, Perez-Rodriguez MM, Basurte-Villamor I, et al. Diagnostic stability and evolution of bipolar disorder in clinical practice: a prospective cohort study. *Acta Psychiatr Scand*. 2007;115(6):473–480.
6. Dassa D, Boyer L, Raymondet P, BottaiT. One or more durations of untreated psychosis? *Acta Psychiatr Scand*. 2011;123(6):494.
7. Fond G, Boyer L, Andrianarisoa M. Risk factors for increased duration of untreated psychosis. Results from the FACE-SZ dataset. *Schizophr Res*. Epub 2017 Sep 6.
8. Hill M, Crumlish N, Clarke M, et al. Prospective relationship of duration of untreated psychosis to psychopathology and functional outcome over 12 years. *Schizophr Res*. 2012;141(2–3):215–221.
9. Nielssen O, Mcgorry P, Castle D, Galletly C. The RANZCP guidelines for Schizophrenia: why is our practice so far short of our recommendations, and what can we do about it? *Aust N Z J Psychiatry*. 2017;51(7):670–674.
10. Coldefy M, Le Neindre C. Les disparités territoriales d'offre et d'organisation des soins en psychiatrie en France: d'une vision segmentée à une approche systémique. Paris: Institut de Recherche et Documentation en Economie de la Santé (IRDES); 2014. Report no 558.
11. Fond G, Boyer L, Boucekine M, et al. Validation study of the Medication Adherence Rating Scale. Results from the FACE-SZ national dataset. *Schizophr Res*. 2017;182:84–89.
12. Caqueo-Urízar A, Urzúa A, Fond G, Boyer L. Medication nonadherence among South American patients with schizophrenia. *Patient Prefer Adherence*. 2017;11:1737–1744.
13. Corréard N, Consoloni JL, Raust A, et al. Neuropsychological functioning, age, and medication adherence in bipolar disorder. *PLoS One*. 2017;12(9):e0184313.
14. Schürhoff F, Fond G, Berna F, et al. A National network of schizophrenia expert centres: an innovative tool to bridge the research-practice gap. *Eur Psychiatry*. 2015;30(6):728–735.
15. Henry C, Etain B, Mathieu F, et al. A French network of bipolar expert centres: a model to close the gap between evidence-based medicine and routine practice. *J Affect Disord*. 2011;131(1–3):358–363.
16. Gaebel W, Janssen B, Zielasek J. Mental health quality, outcome measurement, and improvement in Germany. *Curr Opin Psychiatry*. 2009;22(6):636–642.
17. Herbstman BJ, Pincus HA. Measuring mental healthcare quality in the United States: a review of initiatives. *Curr Opin Psychiatry*. 2009;22(6):623–630.
18. Glied SA, Stein BD, Mcguire TG, et al. Measuring performance in psychiatry: a call to action. *Psychiatr Serv*. 2015;66(8):872–878.
19. Wang DE, Tsugawa Y, Figueroa JF, Jha AK. Association between the centers for medicare and medicaid services hospital star rating and patient outcomes. *JAMA Intern Med*. 2016;176(6):848–850.
20. Trzeciak S, Gaughan JP, Bosire J, Mazzarelli AJ. Association between medicare summary star ratings for patient experience and clinical outcomes in US Hospitals. *J Patient Exp*. 2016;3(1):6–9.
21. Coulter A. Measuring what matters to patients. *BMJ*. 2017;356:j816.
22. Organisation for Economic Co-operation and Development. Recommandations to OECD ministers of health from the high level reflection group on the future of health statistics; 2017. Available from: https://www.oecd.org/els/health-systems/Recommendations-from-high-level-reflection-group-on-the-future-of-health-statistics.pdf. Accessed March 15, 2018.
23. Organisation for Economic Co-operation and Development. OECD Health ministerial statement – The next generation of health reforms; 2017. Available from: https://www.oecd.org/health/ministerial-statement-2017.pdf. Accessed March 15, 2018.

24. Barker DA, Shergill SS, Higginson I, Orrell MW. Patients' views towards care received from psychiatrists. *Br J Psychiatry*. 1996;168(5):641–646.
25. Eriksson KI, Westrin CG. Coercive measures in psychiatric care. Reports and reactions of patients and other people involved. *Acta Psychiatr Scand*. 1995;92(3):225–230.
26. Tessier A, Boyer L, Husky M, Baylé F, Llorca PM, Misdrahi D. Medication adherence in schizophrenia: the role of insight, therapeutic alliance and perceived trauma associated with psychiatric care. *Psychiatry Res*. 2017;257:315–321.
27. Zendjidjian XY, Auquier P, Lançon C, et al. Determinants of patient satisfaction with hospital health care in psychiatry: results based on the SATISPSY-22 questionnaire. *Patient Prefer Adherence*. 2014;8:1457–1464.
28. Zendjidjian XY, Baumstarck K, Auquier P, Loundou A, Lançon C, Boyer L. Satisfaction of hospitalized psychiatry patients: why should clinicians care? *Patient Prefer Adherence*. 2014;8:575–583.
29. Zimmerman M, Gazarian D, Multach M, et al. A clinically useful self-report measure of psychiatric patients' satisfaction with the initial evaluation. *Psychiatry Res*. 2017;252:38–44.
30. Clough BA, Nazareth SM, Casey LM. The therapy attitudes and process questionnaire: a brief measure of factors related to psychotherapy appointment attendance. *Patient*. 2017;10(2):237–250.
31. Mayston R, Habtamu K, Medhin G, et al. Developing a measure of mental health service satisfaction for use in low income countries: a mixed methods study. *BMC Health Services Research*. 2017;17(1):183.
32. Eton DT, Yost KJ, Lai JS, et al. Development and validation of the Patient Experience with Treatment and Self-management (PETS): a patient-reported measure of treatment burden. *Qual Life Res*. 2017;26(2):489–503.
33. Williams AM, Lester L, Bulsara C, et al. Patient Evaluation of Emotional Comfort Experienced (PEECE): developing and testing a measurement instrument. *BMJ Open*. 2017;7(1):e012999.
34. Ortiz G, Schacht L. Psychometric evaluation of an inpatient consumer survey measuring satisfaction with psychiatric care. *Patient*. 2012;5(3):163–173.
35. Olsen RV, Garratt AM, Iversen HH, Bjertnaes OA. Rasch analysis of the Psychiatric Out-Patient Experiences Questionnaire (POPEQ). *BMC Health Serv Res*. 2010;10(1):282.
36. Ruggeri M, Lasalvia A, dall'agnola R. Development, internal consistency and reliability of the Verona service satisfaction scale-European version. EPSILON Study 7. European psychiatric services: inputs linked to outcome domains and needs. *Br J Psychiatry Suppl*. 2000;39:S41–S48.
37. Nordon C, Falissard B, Gerard S, et al. Patient satisfaction with psychotropic drugs: Validation of the PAtient SAtisfaction with Psychotropic (PASAP) scale in patients with bipolar disorder. *Eur Psychiatry*. 2014;29(3):183–190.
38. Ul-Haq I. Patients' Satisfaction with a Psychiatric Day Hospital in the West Galway Catchments Area. *Ir J Psychol Med*. 2012;29(2):85–90.
39. Greenhalgh J, Long AF, Flynn R. The use of patient reported outcome measures in routine clinical practice: lack of impact or lack of theory? *Soc Sci Med*. 2005;60(4):833–843.
40. Hoertel N, Limosin F, Leleu H. Poor longitudinal continuity of care is associated with an increased mortality rate among patients with mental disorders: results from the French National Health Insurance Reimbursement Database. *Eur Psychiatry*. 2014;29(6):358–364.
41. Filipovic-Pierucci A, Samson S, Fagot J-P, Fagot-Campagna A. Estimating the prevalence of depression associated with healthcare use in France using administrative databases. *BMC Psychiatry*. 2017;17(1):1.
42. Verdoux H, Pambrun E, Cortaredona S, et al. Geographical disparities in prescription practices of lithium and clozapine: a community-based study. *Acta Psychiatr Scand*. 2016;133(6):470–480.
43. Tournier M, Cougnard A, Boutouaba-Combe S, Verdoux H. Duration of antidepressant drug treatment and its determinants in France. *Encephale*. 2011;37(Suppl 1):S36–S41.
44. Quantin C, Collin C, Frérot M, et al. Study of algorithms to identify schizophrenia in the SNIIRAM database conducted by the REDSIAM network. *Rev Epidemiol Sante Publique*. 2017;65(Suppl 4):S226–S235.
45. Bezin J, Duong M, Lassalle R, et al. The national healthcare system claims databases in France, SNIIRAM and EGB: powerful tools for pharmacoepidemiology. *Pharmacoepidemiol Drug Saf*. 2017;26(8):954–962.
46. santé H.Ade. Fiche descriptive de l'indicateur de qualité et de sécurité des soins: Tenue du dossier patient (TDP) en santé mentale; 2015. Available from: https://www.has-sante.fr/portail/upload/docs/application/pdf/2010-12/ipaqss_psy_fiche_tdp.pdf. Accessed January 28, 2018.
47. Boyer L, Dassa D, Samuelian JC, Lancon C. Interrogations autour de la valorisation de l'activité en psychiatrie [Questions about the future French mental health finance]. *L'Encéphale*. 2011;37(1):1–3. French.
48. Boyer L, Fond G, Devictor B, et al. Réflexion sur la péréquation financière de psychiatrie en France [Reflection on the psychiatric financial allocation in France]. *L'Encéphale*. 2016;42(4):379–381. French.
49. Michel P, Baumstarck K, Lancon C, et al. Modernizing quality of life assessment: development of a multidimensional computerized adaptive questionnaire for patients with schizophrenia. *Qual Life Res*. 2018;27(4):1041–1054.
50. Gibbons RD, Weiss DJ, Kupfer DJ, et al. Using computerized adaptive testing to reduce the burden of mental health assessment. *Psychiatr Serv*. 2008;59(4):361–368.
51. Gibbons RD, Weiss DJ, Pilkonis PA, et al. Development of the CAT-ANX: a computerized adaptive test for anxiety. *Am J Psychiatry*. 2014;171(2):187–194.
52. Gibbons RD, Weiss DJ, Pilkonis PA, et al. Development of a computerized adaptive test for depression. *Arch Gen Psychiatry*. 2012;69(11):1104–1112.
53. Kingsley C, Patel S. Patient-reported outcome measures and patient-reported experience measures. *BJA Education*. 2017;17(4):137–144.
54. Greene SM, Tuzzio L, Cherkin D. A framework for making patient-centered care front and center. *Perm J*. 2012;16(3):49–53.
55. Miller D, Steele Gray C, Kuluski K, Cott C. Patient-Centered Care and Patient-Reported Measures: let's look before we leap. *Patient*. 2015;8(4):293–299.
56. Weiss DJ. Computerized adaptive testing for effective and efficient measurement in counseling and education. *Meas Eval Couns Dev*. 2004;37(2):70–84.
57. Reeve BB, Hays RD, Bjorner JB, et al. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Med Care*. 2007;45(5 Suppl 1):S22–S31.
58. Rose M, Bjorner JB, Becker J, Fries JF, Ware JE. Evaluation of a preliminary physical function item bank supported the expected advantages of the Patient-Reported Outcomes Measurement Information System (PROMIS). *J Clin Epidemiol*. 2008;61(1):17–33.
59. Cella D, Riley W, Stone A, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *J Clin Epidemiol*. 2010;63(11):1179–1194.
60. DeWalt DA, Rothrock N, Yount S, Stone AA; PROMIS Cooperative Group. Evaluation of item candidates: the PROMIS qualitative item review. *Med Care*. 2007;45(5 Suppl 1):S12–S21.
61. Yrondi A, Bennabi D, Haffen E, et al. Significant Need for a French Network of Expert Centers Enabling a Better Characterization and Management of Treatment-Resistant Depression (Fondation FondaMental). *Front Psychiatry*. 2017;8:244.
62. Acquadro C, Conway K, Hareendran A, Aaronson N; European Regulatory Issues and Quality of Life Assessment (ERIQA) Group. Literature review of methods to translate health-related quality of life questionnaires for use in multinational clinical trials. *Value Health*. 2008;11(3):509–521.
63. Bullinger M, Alonso J, Apolone G, et al. Translating health status questionnaires and evaluating their quality: the IQOLA Project approach. International Quality of Life Assessment. *J Clin Epidemiol*. 1998;51(11):913–923.
64. Petersen MA, Giesinger JM, Holzner B, et al. Psychometric evaluation of the EORTC computerized adaptive test (CAT) fatigue item pool. *Qual Life Res*. 2013;22(9):2443–2454.

65. Abberger B, Haschke A, Krense C, Wirtz M, Bengel J, Baumeister H. The calibrated, unidimensional anxiety item bank for cardiovascular patients provided the basis for anxiety assessment in cardiovascular rehabilitation patients. *J Clin Epidemiol*. 2013;66(8):919–927.
66. Baumstarck K, Boyer L, Boucekine M, Michel P, Pelletier J, Auquier P. Measuring the quality of life in patients with multiple sclerosis in clinical practice: a necessary challenge. *Mult Scler Int*. 2013;2013(2):1–8.
67. Kaiser HF. An index of factorial simplicity. *Psychometrika*. 1974;39(1): 31–36.
68. Fabrigar LR, Wegener DT, Maccallum RC, Strahan EJ. Evaluating the use of exploratory factor analysis in psychological research. *Psychol Methods*. 1999;4(3):272–299.
69. Reise SP, Morizot J, Hays RD. The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Qual Life Res*. 2007;16(S1):19–31.
70. Reise SP, Moore TM, Haviland MG. Bifactor models and rotations: exploring the extent to which multidimensional data yield univocal scale scores. *J Pers Assess*. 2010;92(6):544–559.
71. Kaiser HF, Caffrey J. Alpha factor analysis. *Psychometrika*. 1965;30(1): 1–14.
72. Cattell RB. The scree test for the number of factors. *Multivariate Behav Res*. 1966;1(2):245–276.
73. Horn JL. A rationale and test for the number of factors in factor analysis. *Psychometrika*. 1965;30(2):179–185.
74. Kline P. *An Easy Guide to Factor Analysis*. London, UK: Routledge; 1994.
75. Nunnaly JC, Bernstein IH. *Psychometric Theory*. 3rd ed. New York, NY: McGraw-Hill; 1994.
76. Hooper D, Coughlan J, Mullen MR. Structural equation modelling: guidelines for determining model fit. *Electron J Bus Res Methods*. 2008;6(1):53–60.
77. Kline RB. *Principles and Practice of Structural Equation Modeling*. 3rd ed. New York, NY: Guilford Press; 2011.
78. Hu Li-tze, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Modeling*. 1999;6(1):1–55.
79. Bentler PM. Comparative fit indexes in structural models. *Psychol Bull*. 1990;107(2):238–246.
80. Embretson SE, Reise SP. *Item Response Theory for Psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates; 2000.
81. Bjorner JB, Kosinski M, Ware JE. Calibration of an item pool for assessing the burden of headaches: an application of item response theory to the headache impact test (HIT). *Qual Life Res*. 2003;12(8): 913–933.
82. Fliege H, Becker J, Walter OB, Bjorner JB, Klapp BF, Rose M. Development of a computer-adaptive test for depression (D-CAT). *Qual Life Res*. 2005;14(10):2277–2291.
83. Michel P, Auquier P, Baumstarck K, et al. Development of a cross-cultural item bank for measuring quality of life related to mental health in multiple sclerosis patients. *Qual Life Res*. 2015;24(9):2261–2271.
84. Bjorner JB, Chang CH, Thissen D, Reeve BB. Developing tailored instruments: item banking and computerized adaptive assessment. *Qual Life Res*. 2007;16(S1):95–108.
85. Muraki E. A generalized partial credit model. In: van der Linden WJ, Hambleton RK, editors. *Handbook of Modern Item Response Theory*. New York, NY: Springer; 1997:153–164.
86. Bock RD, Aitkin M. Marginal maximum likelihood estimation of item parameters: application of an EM algorithm. *Psychometrika*. 1981; 46(4):443–459.
87. Rose M, Bjorner JB, Becker J, Fries JF, Ware JE. Evaluation of a preliminary physical function item bank supported the expected advantages of the Patient-Reported Outcomes Measurement Information System (PROMIS). *J Clin Epidemiol*. 2008;61(1):17–33.
88. Bjorner JB, Kosinski M, Ware Jr JE. The feasibility of applying item response theory to measures of migraine impact: a re-analysis of three clinical studies. *Qual Life Res*. 2003;12(8):887–902.
89. Samejima F. Graded response model. In: van der Linden WJ, Hambleton RK, editors. *Handbook of Modern Item Response Theory*. New York, NY: Springer; 1997:85–100.
90. Maydeu-Olivares A, Drasgow F, Mead AD. Distinguishing among paranletric item response models for polychotomous ordered data. *Appl Psychol Meas*. 1994;18(3):245–256.
91. Cook KF, Dodd BG, Fitzpatrick SJ. A comparison of three polytomous item response theory models in the context of testlet scoring. *J Outcome Meas*. 1999;3(1):1–20.
92. Wright BD, Linacre JM. Reasonable mean-square fit values. *Rasch Meas Trans*. 1994;8:370.
93. Rogers HJ. Differential item functioning. In: Everitt BS, Howell DC, editors. *Encyclopedia of Statistics in Behavioral Science*. Chichester, UK: John Wiley & Sons, Ltd; 2005:485–490.
94. Zieky M. Differential item functioning. In: Holland PW, Wainer H, editors. *Pratical Questions in the Use of DIF Statistics in Test Development*. Hillsdale, NJ: Lawrence Erlbaum Associates; 1993:337–347.
95. Choi SW, Gibbons LE, Crane PK. Lordif: an R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and monte carlo simulations. *J Stat Softw*. 2011;39(8):1–30.
96. Zumbo BD. *A Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-Type (Ordinal) Item Scores*. Ottawan, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense; 1999.
97. Bock RD, Mislevy RJ. Adaptive EAP estimation of ability in a microcomputer environment. *Appl Psychol Meas*. 1982;6(4):431–444.
98. Choi SW, Swartz RJ. Comparison of CAT item selection criteria for polytomous items. *Appl Psychol Meas*. 2009;33(6):419–440.
99. Harvill LM. Standard error of measurement. *Educ Meas Issues Pract*. 1991;10(2):33–41.
100. Michel P, Baumstarck K, Ghattas B, et al. A multidimensional computerized adaptive short-form quality of life questionnaire developed and validated for multiple sclerosis: The MusiQoL-MCAT. *Medicine*. 2016;95(14):e3068.
101. Wainer H, Dorans NJ, Eignor D, Flaugher R, Green BF, Mislevy RJ. *Computerized Adaptive Testing: A Primer*. 2nd ed. Mahwah, NJ: Lawrence Erlbaum Associates; 2000.
102. Michel P, Baumstarck K, Loundou A, Ghattas B, Auquier P, Boyer L. Computerized adaptive testing with decision regression trees: an alternative to item response theory for quality of life measurement in multiple sclerosis. *Patient Prefer Adherence*. 2018;12:1043–1053.
103. Sylvia LG, Hay A, Ostacher MJ, et al. Association between therapeutic alliance, care satisfaction, and pharmacological adherence in bipolar disorder. *J Clin Psychopharmacol*. 2013;33(3):343–350.
104. Misdrahi D, Petit M, Blanc O, Bayle F, Llorca PM. The influence of therapeutic alliance and insight on medication adherence in schizophrenia. *Nord J Psychiatry*. 2012;66(1):49–54.
105. Cook EL, Harman JS. A comparison of health-related quality of life for individuals with mental health disorders and common chronic medical conditions. *Public Health Rep*. 2008;123(1):45–51.
106. Prigent A, Auraaen A, Kamendje-Tchokobou B, Durand-Zaleski I, Chevreul K. Health-related quality of life and utility scores in people with mental disorders: a comparison with the non-mentally ill general population. *Int J Environ Res Public Health*. 2014;11(3):2804–2817.
107. Wilson KA, Dowling AJ, Abdolell M, Tannock IF. Perception of quality of life by patients, partners and treating physicians. *Qual Life Res*. 2000;9(9):1041–1052.
108. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA*. 2013;310(20):2191–2194.
109. Smith PC, Mossialos E, Papanicolas I, Leatherman S, editors. *Performance Measurement for Health System Improvement: Experiences, Challenges and Prospects*. Cambridge, NY: Cambridge University Press; 2009.

110. Killaspy H, Cardoso G, White S, et al. Quality of care and its determinants in longer term mental health facilities across Europe: a cross-sectional analysis. *BMC Psychiatry*. 2016;16(1):31.
111. Killaspy H, White S, Dowling S, et al. Adaptation of the Quality Indicator for Rehabilitative Care (QuIRC) for use in mental health supported accommodation services (QuIRC-SA). *BMC Psychiatry*. 2016;16(1):101.
112. Killaspy H, King M, Wright C, et al. Study protocol for the development of a European measure of best practice for people with long term mental health problems in institutional care (DEMoBinc). *BMC Psychiatry*. 2009;9(1):36.
113. Bramesfeld A, Amaddeo F, Caldas-de-Almeida J, et al. Monitoring mental healthcare on a system level: country profiles and status from EU countries. *Health Policy*. 2016;120(6):706–717.
114. Gutiérrez-Colosía MR, Salvador-Carulla L, Salinas-Pérez JA, et al. Standard comparison of local mental health care systems in eight European countries. *Epidemiol Psychiatr Sci*. 2017;15:1–14.
115. Cella D, Yount S, Rothrock N, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. *Med Care*. 2007;45(5 Suppl 1):S3–S11.
116. Kelley TA. International Consortium for Health Outcomes Measurement (ICHOM). *Trials*. 2015;16(S3):O4.
117. Coldefy M. The evolution of psychiatric care systems in Germany, England, France and Italy: similarities and differences. *Health Economics*. 2012;180:1–8.
118. Baumstarck K, Boyer L, Boucekine M, et al. Self-reported quality of life measure is reliable and valid in adult patients suffering from schizophrenia with executive impairment. *Schizophr Res*. 2013;147(1):58–67.
119. Lagha E, Noble A, Smith A, Denvir MA, Leslie SJ. Patient Reported Experience Measures (PREMs) in chronic heart failure. *J R Coll Physicians Edinb*. 2012;42(4):301–305.
120. Saunders CL, Elliott MN, Lyratzopoulos G, Abel GA. Do differential response rates to patient surveys between organizations lead to unfair performance comparisons? Evidence from the English Cancer Patient Experience Survey. *Med Care*. 2016;54(1):45–54.
121. Ahmed F, Burt J, Roland M. Measuring patient experience: concepts and methods. *Patient*. 2014;7(3):235–241.
122. Eisen SV, Shaul JA, Leff HS, Stringfellow V, Clarridge BR, Cleary PD. Toward a national consumer survey: evaluation of the CABHS and MHSIP instruments. *J Behav Health Serv Res*. 2001;28(3):347–369.
123. Guy W. Clinical Global Impressions. ECDEU Assessment Manual for Psychopharmacology. Rockville, MD: US Department of Health, Education, and Welfare Public Health Service Alcohol, Drug Abuse, and Mental Health Administration; 1976.
124. Berk M, Ng F, Dodd S, et al. The validity of the CGI severity and improvement scales as measures of clinical effectiveness suitable for routine clinical use. *J Eval Clin Pract*. 2008;14(6):979–983.
125. Endicott J, Spitzer R, Fleiss J, Cohen J. The global assessment scale. A procedure for measuring overall severity of psychiatric disturbance. *Arch Gen Psychiatry*. 1976;33(6):766–771.
126. Attkisson CC, Zwick R. The client satisfaction questionnaire. Psychometric properties and correlations with service utilization and psychotherapy outcome. *Eval Program Plann*. 1982;5(3):233–237.
127. Misdrahi D, Verdoux H, Lançon C, Bayle F. The 4-Point ordinal Alliance Self-report: a self-report questionnaire for assessing therapeutic relationships in routine mental health. *Compr Psychiatry*. 2009;50(2):181–185.
128. Fialko L, Garety PA, Kuipers E, et al. A large-scale validation study of the Medication Adherence Rating Scale (MARS). *Schizophr Res*. 2008;100(1–3):53–59.
129. Leplege A, Ecosse E, Verdier A, Perneger TV. The French SF-36 Health Survey: translation, cultural adaptation and preliminary psychometric evaluation. *J Clin Epidemiol*. 1998;51(11):1013–1023.
130. Morse JM. The significance of saturation. *Qual Health Res*. 1995;5(2):147–149.
131. Mason M. Sample size and saturation in PhD studies using qualitative interviews. *Forum Qual Soc Res*. 2010;11(3):Art 8.
132. Ritchie J, Lewis J, Elam G. Designing and selecting samples. In: Ritchie J, Lewis J, editors. *Qualitative Research Practice: A Guide for Social Science Students and Researchers*. Thousand Oaks, CA: Sage Publications; 2003:77–108.
133. Edelen MO, Reeve BB. Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Qual Life Res*. 2007;16(S1):5–18.
134. Nguyen TH, Han HR, Kim MT, Chan KS. An introduction to item response theory for patient-reported outcome measurement. *Patient*. 2014;7(1):23–35.
135. Tsutakawa RK, Johnson JC. The effect of uncertainty of item parameter estimation on ability estimates. *Psychometrika*. 1990;55(2):371–390.
136. Cappelleri JC, Jason Lundy J, Hays RD. Overview of classical test theory and item response theory for the quantitative assessment of items in developing patient-reported outcomes measures. *Clin Ther*. 2014;36(5):648–662.
137. Anthoine E, Moret L, Regnault A, Sébille V, Hardouin JB. Sample size used to validate a scale: a review of publications on newly-developed patient reported outcomes measures. *Health Qual Life Outcomes*. 2014;12(1):176.