Article:
McKenzie, E., Matkin, L., Sousa Fialho, L. et al. (23 more authors) (2022) Developing an international standard set of patient-reported outcome measures for psychotic disorders. Psychiatric Services, 73 (3). pp. 249-258. ISSN 1075-2730

https://doi.org/10.1176/appi.ps.202000888

Title:

Developing an International Standard Set of Patient-Reported Outcome Measures for Psychotic Disorders

Disclosures and acknowledgments:

Dr. X has been a consultant and/or advisor to or have received honoraria from: Acadia, Alkermes, Allergan, Angelini, Axsome, Gedeon Richter, Gerson Lehrman Group, Indivior, IntraCellular Therapies, Janssen/J&J, Karuna, LB Pharma, Lundbeck, MedAvante-ProPhase, MedInCell, Medscape, Merck, Mylan, Neurocrine, Noven, Otsuka, Pfizer, Recordati, Rovi, Servier, Sumitomo Dainippon, Sunovion, Supernus, Takeda, and Teva. Dr. X provided expert testimony for Janssen and Otsuka. Dr. X served on a Data Safety Monitoring Board for Lundbeck, Rovi, Supernus, and Teva. Dr. X has received grant support from Janssen and Takeda and is also a stock option holder of LB Pharma.

Word count: 2983

Highlights:

- Patient-reported outcomes reflect the outcomes important to patients and can be used for comparisons across programs and countries.
- An international working group used systematic searches to identify and assess the quality of psychometric evidence supporting patient-reported outcome measures for psychosis and associated risk adjustment factors.
- Service users, clinicians and researchers were involved in a consensus process to reduce the measures to a standard set which can be completed in about 35 minutes.
- The outcomes are assessed by nine measures which cover symptoms, recovery, functioning and treatment.

Abstract

Objective: The objective of this project was to develop a set of patient-reported outcome measures for adolescents and adults meeting criteria for a psychotic disorder.

Methods: A research team and an international consensus working group including service users, clinicians, and researchers worked together in an iterative process using a modified Delphi consensus technique which included video-conference calls, online surveys and focus groups. The research team conducted systematic literature searches to identify outcomes, outcome measures, and risk adjustment factors. After identifying outcomes important to service users, the consensus working group selected outcome measures, risk adjustment factors and the final set of outcome measures. International stakeholder groups consisting of over 100 professionals and service users reviewed and commented on the final set.

Results: The consensus working group identified four outcome domains; symptoms, personal recovery, functioning and treatment. The domains encompassed 14 outcomes of importance to service users. The research team identified 131 measures from the literature. The consensus working group selected nine measures in an outcome set that takes approximately 35 minutes to complete.

Conclusions: A set of patient-reported outcome measures for use in routine clinical practice was identified. The set is free to the user, available in at least two languages and reflects outcomes important to service users. Clinicians can use the set to improve clinical decision making, and administrators and researchers can use it to learn from comparing program outcomes.

Introduction

A patient-reported outcome (PRO), as measured by patient-reported outcome measures (PROMs), is any aspect of a patient's health status that comes directly from the patient (1). PROMs can be used to improve clinical care (2, 3), inform clinicians and patients about treatment progress, create, compare, and aggregate scores at a high level to inform policy, or for approval of drugs and devices (4, 5). Research and application of PROMs in health care, particularly in the management of chronic disorders, has increased over the last twenty years (6). The use of PROMs is a focus of patient-centred care (7). Notwithstanding the challenges to implementing large-scale PRO systems in health care, factors that increase the rate of PROM collection include training providers on the use of PROMs, implementing software to register and work with PROMs in daily practice, administrative surveillance of collection rates, and the presence of local clinical champions (3). International application of PROMs requires high standards for translation (8).

We identified two examples of large-scale implementation of recommended PROMs to mental health services in specific programs in the UK (9) and in routine outcome measurement in Australia (10). However, a Cochrane review of routine monitoring using PROMs for improving treatment of common mental health disorders in adults found insufficient evidence to support routine monitoring and identified the need for “more research of better quality”, including measuring a range of relevant outcomes (11).

The outcomes important to patients can be identified through focus groups, in-depth interviews, target population surveys, qualitative synthesis of the literature and content

analysis of available data sources (12). Ideally, PROMs display strong psychometric properties including a conceptual and measurement model, reliability, validity, responsiveness, interpretability, alternative modes of administration, and cross-cultural and linguistic adaptations (13, 14). Practical implementation of measures warrants consideration, and includes identifying the goals for collecting PROs, selection of patients, and the setting and timing of assessments (15).

Psychotic disorders, including schizophrenia and bipolar disorder, are responsible for significant burden to service users, families, and health systems worldwide (16). Although schizophrenia is a low prevalence disorder with an estimated population prevalence at less than 1% (17), it is associated with adverse mental and physical health outcomes, a high degree of disability, high healthcare costs, and a 15-year reduction in life expectancy (18-23). Bipolar disorder has a slightly higher prevalence at approximately 1%, with psychiatric as well as physical health burden resulting from its early onset, severity, and chronicity (21-25). While clinical recovery relates to the remission of symptoms and the return of functioning (26) the meaning of personal recovery to service users is broader and a process that encompasses many aspects of life and promotion of an individual's strength and potential (27). There is evidence that supports recovery as a feasible outcome for schizophrenia-spectrum disorders and bipolar disorder (28, 29). At the health systems level, there has been a shift in focus toward a recovery orientation and personalized care (30, 31). The broad impact of psychotic disorders has spurred investigators to examine a range of PROs in schizophrenia, and patient-reported quality of life (32) and functional outcome (33) in bipolar disorder.

Through use of a consensus building process, the International Consortium for Health Outcomes Measurement (ICHOM) developed and implemented standard sets of patient outcomes for use in routine clinical practice for various medical conditions. The process is supported by identification of the evidence, and systematic, critical evaluation of measures and their psychometric properties. ICHOM organized a working group of psychosis experts, including clinicians, researchers, and service users, to identify a set of PROMs to monitor individual treatment outcomes or to compare outcomes of similar mental health services, with a view to establish the value of these services. The value of health care can be defined as the patient outcomes relative to the costs for obtaining those outcomes (34). Outcome assessment can be guided by a set of PROs for a specific disorder, as exemplified by the set for depression and anxiety developed by ICHOM (35). The specific aims of this study were to develop a standard minimum set of PROMs for psychotic disorders that can be used anywhere in the world, and to identify a set of risk adjustment factors to enhance utility of comparisons across treatment modalities, institutions, and systems.

Methods

The research team included a chair, project manager, research fellow, and five research associates. The working group members (n=19) included service users (n=3) and were selected to represent diverse professions and geographic areas. Ten areas of expertise were represented:: psychometrists, psychiatrists, mental health nurses, psychologists, health economists, epidemiologists, national clinical quality programs, health service researchers, social workers, and service users, with membership from 11 countries (Australia, Canada,

Denmark, Greece, India, Israel, Japan, Mexico, Netherlands, the United Kingdom, and the United States).

Service User Focus Groups

Prior to commencement of the working group meetings, two focus groups were held with four service users to identify outcomes of highest importance to them. Service user recruitment occurred through recommendations from patient organizations and from working group members.

Systematic Literature Review to Identify Outcomes

Between January and March 2019, systematic literature searches were conducted to identify outcomes related to schizophrenia and bipolar type I in adolescent and adult populations. The databases Medline, PubMed and PsycInfo were searched for publications between January 2013 and January 2019. Search strategies are provided in the online supplement.

The Cochrane Highly Sensitive Search Strategy (CHSSS) for identifying randomized trials (36) was first used to identify outcomes for schizophrenia-spectrum and bipolar I disorder. Additional searches that excluded randomized trials were conducted in PsycInfo and Medline. This included qualitative research that examined service users' perspectives on outcomes of importance and the impact of schizophrenia or bipolar disorder on service users' lives. Supplementary sources from working group members and reference lists in identified papers were additional sources of PROs and PROMs.

Consensus Process

A modified Delphi consensus process was used to select the outcome set (37). The process involved reaching consensus in five main areas: 1. Scope, to determine which psychotic disorders, patient populations, and treatments; 2. Outcomes, to identify a minimum number of outcomes for inclusion in the set; 3. Measures to assess each outcome; 4. Definitions and time points for outcomes assessment; and 5: Risk adjustment factors to enable comparison between providers implementing the set. The research team prepared and distributed presentations for review prior to each videoconference. The five main areas were discussed during the videoconferences and feedback from the working group was incorporated. Members rated and provided feedback on each item under review in the five main areas, in online surveys.

Following ICHOM processes outlined a priori, inclusion in the set required a minimum of 80% of the working group voting an item as "essential" (score of 7-9) in the first or second Delphi round. When consensus was not reached by voting, the item was discussed and revisited in the next videoconference and survey. Outcomes were excluded if a minimum of 80% of the working group voted an item as "not recommended" (score 1-3). The working group voted on all inconclusive outcomes in the third and final survey round, in which response options were "include" or "exclude", and inclusion required only 70% consensus. The development process is illustrated in Figure 1.

Identification of Potential Outcome Measures from the Systematic Literature Review

Publications identified in the systematic literature review were the source of potential outcome measures. After development of a comprehensive measures list, a search filter was used to facilitate measure selection (38) that identified studies in PubMed reporting psychometric

properties of measures. The research team excluded measures that had a cost associated with use.

Assessment of PROMs

The COnsensus-based Standards for the selection of health outcomes Measurement Instruments (COSMIN) checklist was used to assess the psychometric properties of measures, including reliability (test-retest, internal consistency), validity (content validity, face validity, and construct validity) and responsiveness (sensitivity to change) (14). In addition, the working group considered the feasibility of collecting the measures, including length of time to administer, training needed, and international applicability.

Breakout Sessions to Narrow the List of Measures

After the COSMIN checklist was applied, four breakout sessions were held to review and reduce the number of potential measures. The sessions were held with a small number of working group members with lived experience and/or professional expertise in the areas of functioning, personal recovery, symptoms, and treatment. Participants narrowed the list of potential measures and established a proposal for the wider working group to discuss and endorse. Any measures that took longer than 20 minutes to complete were removed.

Risk Adjustment Factors

A preliminary list of risk adjustment factors and definitions was developed based on risk adjustment factors identified from the systematic literature searches, national registries, and review of existing ICHOM standard sets. Factors were identified based on evidence of their

effect on the outcomes selected. Demographic and clinical factors were assessed on relevance and feasibility of measurement. Further, to harmonize across ICHOM mental health standard sets, ICHOM mental health working group chairs developed a list of factors to reduce burden of implementation in service users with more than one diagnosis. The harmonized list was presented to each mental health consensus working group. Factors voted for inclusion reached 70% consensus.

Open Review and Patient Validation

After development, the set was distributed to international stakeholder groups including professionals, adult service users, and carers in two separate stakeholder surveys to obtain feedback on outcomes, measures, risk adjustment factors and timing of outcome collection. Respondents were recruited via networks identified by the research team and working group members through email and social media, national and international patient organizations. Service users and carers were asked to rate the importance of each outcome using a 9-point Likert scale and were provided space to suggest additional outcomes. No institutional review board (IRB) approval or informed consent was necessary for this study.

Results

Scope

The working group selected both schizophrenia-spectrum disorders (as classified in ICD-11 and DSM-5) and bipolar disorder type I (as classified in ICD-11). The set is limited to the adolescent (age ≥12 years) and adult populations.

Service User Focus Groups

The core outcomes identified as important to service users included an improvement in positive and negative symptoms, physical health, medication side effects, personal recovery, and prevention of relapse.

Identifying PROs

The literature searches identified a total of 9,118 references. Of those, n=758 were eligible for full-text review (diagram available online). Supplemental sources were recommended by working group members. After removing duplicates and measures with associated costs and in languages other than English, a total of 131 measures were examined that assess symptoms, personal recovery, functioning and treatment in psychosis (online supplement).

Domain, Outcome and Measure Selection by Consensus Working Group

The working group identified four outcome domains (symptoms, recovery, functioning, and treatment) that encompassed 14 outcomes (see Table 1). Symptoms included depressive symptoms, suicidal ideation and behavior, positive symptoms, negative symptoms (schizophrenia), mania/hypomania (bipolar I), sleep quality and relapse rate as measured by hospitalizations. Personal recovery included quality of life measures. Functioning included global, social and role functioning, in addition to physical health. Medication side effects were included under treatment outcomes. Each core outcome identified by service users in the focus groups was included in the final set.

Fifty-seven measures were presented to the working group for vote (online supplement). Outcome measures were reviewed in their entirety, with consideration given to psychometric

properties, previous use in the specified population, the number and wording of items, and time taken to complete (see Table 2). Response rates were 80%, 85%, 80%, and 80% for the first through fourth modified Delphi processes, respectively.

Service users provided feedback regarding item wording and appropriateness, and their opinion about the ability of the measure to capture the outcome for individuals with a psychotic disorder. This feedback was summarized in tables and presented alongside other working group member feedback. One hundred and forty-seven measures were initially mapped to the 14 outcomes, and 39 presented via shortlists in breakout sessions.

Evaluation of Measures

Measures were assessed on psychometric properties, including acceptable interrater reliability ($r \geq 0.70$)(39), internal consistency (Cronbach's $\alpha \geq 0.70$)(40), and evidence of sensitivity to change reflected by change in scores measured over time, consistent with a priori hypotheses about anticipated treatment outcome (Table 2). Measures rated as strong demonstrated acceptable interrater reliability, internal consistency, and evidence of sensitivity to change. Measures rated as mixed had only a single evaluation identified or mixed evidence from several sources or strong evidence only for certain items. Measures rated as weak demonstrated below threshold evidence.

The working group selected nine measures in an outcome set that takes approximately 35 minutes to complete. The selected measures are freely available to the user in English language.

Timepoint Recommendations

The outcome assessment timeline was proposed by the working group to best achieve a balance between the clinically relevant times when outcomes may be expected to change, and pragmatic concerns in data collection (41). The working group recommended assessment of outcomes prior to treatment as a baseline, and then every six months, and risk adjustment factors assessed at baseline and annually thereafter.

Risk Adjustment Factors

Harmonization of risk adjustment factors across mental health sets resulted in the addition of two factors to the preliminary list of risk adjustment factors; trauma, as assessed by adverse childhood experiences, and contact with law enforcement (see Table 3).

Open Review and Patient Validation

Ninety-five professionals living in Australia, Canada, Chile, Nigeria, Sweden, the UK, and the USA responded to the open review survey. Service users and carers (n=25) were from Australia, the UK and USA.

Overall endorsement for the set and its elements exceeded the required 70%. Of the professional experience participants, support for outcome domains ranged between 77-93%, and for outcome measures, between 61-84%. Of the lived experience participants, 92% agreed the measures are useful to collect and 91% stated the list captured all important outcomes. Endorsement of included outcomes ranged between 72-92%.

Discussion:

The research team identified numerous PROMs for both schizophrenia-spectrum disorders and bipolar disorder. The process involved a review to identify the measures and a subsequent review to assess the measures' psychometric properties (42), a similar process to other reviewers of PROMs. The consensus process was successful in reducing the number of measures to a pragmatic set for use in routine practice. Service users were integral to the development of the set, from initial identification of core outcomes that mattered most to them, to assessment of measures' face validity and comprehensibility of items. The open review and patient validation phase helped ensure the interpretability and cultural sensitivity of the set.

In addition to the well-recognized symptoms associated with psychosis, sleep problems are common in people with schizophrenia-spectrum disorders and bipolar I disorder, negatively impact functioning and well-being, and associated with a reduced ability or opportunity to participate in valued activities (43). Sleep quality, as assessed by the PROMIS-Sleep measure, is included in the set.

The importance of personal recovery to service users was pronounced. The personal recovery measure in the set has good psychometric properties and has been used in published research on mental health populations (44-46). We did not find a positive symptom measure developed and tested exclusively in samples of people with schizophrenia. However, the MCSI has been used in large-scale studies of populations with severe mental illness (47). We did not identify a self-report measure for negative symptoms. Due to its length and the availability of only one language version, the Clinical Assessment Interview for Negative Symptoms: Self report (48), a 30-item measure, is not recommended for inclusion in the outcome set. As a best alternative,

the working group suggested a PROM, the Recovering Quality of Life (ReQol-20) (45). Historically, a clinician-rated outcome measure (CROM), the Quality of Life Scale (49), has been used as a negative symptom measure. There was less research to support decision-making on adolescent measures.

Limitations of this work include a low number of service users and no service users with lived experience of bipolar I disorder in the working group. The inclusion of service users in developing PROMS is important yet remains challenging. Few studies include them at all stages of development (50). In this study, service users were recruited after the design stage and before the decision to include PROMS for bipolar disorder. At project commencement, two service user-only focus groups were held. Additionally, we paid specific attention to recovery measure studies that involved service users in their development and evaluation. A limitation of the final set is redundancy in some items across measures. For example, sleep is assessed in the PROMIS-Sleep measure and in the depression and mania measures. This commonly encountered issue could be addressed in future research using statistical methods to address overlap across the entire outcome set. Consistent with a systematic review of PROMS and CROMS for assessing youth outcomes, we found broader outcome measures developed for adolescents, including measures for Quality of Life. Targeted measures for symptoms and side effects were often developed for adults and rarely tested on, or adapted for, adolescents (51). This highlights the need to validate outcome measures for the adolescent population.

There are important psychometric properties that were not included in our selection criteria. These include meaningful change thresholds (52) and severity thresholds such as mild, moderate and severe which can be linked to treatment decisions (53). These properties were

present to varying degrees in the selected measures yet not used as selection criteria. A critical evaluation of each measure was therefore beyond the scope of this project.

We did not address alternative modes of administration, an important consideration in implementation. However, a meta-analytic review concluded substantial evidence indicating the equivalence of computer- and paper- administered PROs (54).

Conclusions: A standardized set of nine PROs was identified. The set can be used to support measurement-based care and, in combination with risk adjustment factors, to compare program outcomes. Finally, the set can be used to support development of value-based health care for people with psychotic disorders.

References

1. Services DHH: Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance; in Health Qual Life Outcomes, 2006
2. Greenhalgh J: The applications of PROs in clinical practice: what are they, do they work, and why? Qual Life Res 18:115-23, 2009
3. Sisodia RC, Dankers C, Orav J, et al.: Factors Associated With Increased Collection of Patient-Reported Outcomes Within a Large Health Care System. JAMA Netw Open 3:e202764, 2020
4. Acquadro C, Berzon, R., Dubois, D, Leidy, NK, Marquis, P Incorporating the patient's perspective into drug development and communication: an ad hoc Task Force Report of the Patient-Reported Outcomes (PRO) Harmonization Group Meeting at the Food and Drug Administration. Value in Health 6:522-31, 2003
5. Evans JP, Smith A, Gibbons C, et al.: The National Institutes of Health Patient-Reported Outcomes Measurement Information System (PROMIS): a view from the UK. Patient Relat Outcome Meas 9:345-52, 2018
6. 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 63:1179-94, 2010
7. Lavallee DC, Chenok KE, Love RM, et al.: Incorporating Patient-Reported Outcomes Into Health Care To Engage Patients And Enhance Care. Health Aff (Millwood) 35:575-82, 2016
8. Eremenco S, Pease S, Mann S, et al.: Patient-Reported Outcome (PRO) Consortium translation process: consensus development of updated best practices. J Patient Rep Outcomes 2:12, 2017
9. Clark DM, Canvin L, Green J, et al.: Transparency about the outcomes of mental health services (IAPT approach): an analysis of public data. Lancet 391:679-86, 2018
10. Burgess P, Pirkis J, Coombs T: Routine outcome measurement in Australia. Int Rev Psychiatry 27:264-75, 2015
11. Kendrick T, El-Gohary M, Stuart B, et al.: Routine use of patient reported outcome measures (PROMs) for improving treatment of common mental health disorders in adults. Cochrane Database Syst Rev 7:Cd011119, 2016
12. Turner RR, Quittner AL, Parasuraman BM, et al.: Patient-reported outcomes: instrument development and selection issues. Value Health 10 Suppl 2:S86-93, 2007
13. Valderas JM, Ferrer M, Mendívil J, et al.: Development of EMPRO: a tool for the standardized assessment of patient-reported outcome measures. Value Health 11:700-8, 2008
14. Mokkink LB, Prinsen CA, Bouter LM, et al.: The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) and how to select an outcome measurement instrument. Braz J Phys Ther 20:105-13, 2016
15. Snyder CF, Aaronson NK, Choucair AK, et al.: Implementing patient-reported outcomes assessment in clinical practice: a review of the options and considerations. Qual Life Res 21:1305-14, 2012
16. Vigo D, Thornicroft G, Atun R: Estimating the true global burden of mental illness. Lancet Psychiatry 3:171-8, 2016
17. Saha S, Chant D, McGrath J: Meta-analyses of the incidence and prevalence of schizophrenia: conceptual and methodological issues. Int J Methods Psychiatr Res 17:55-61, 2008

18. Charlson FJ, Ferrari AJ, Santomauro DF, et al.: Global Epidemiology and Burden of Schizophrenia: Findings From the Global Burden of Disease Study 2016. Schizophr Bull 44:1195-203, 2018
19. Hjorthøj C, Stürup AE, McGrath JJ, et al.: Years of potential life lost and life expectancy in schizophrenia: a systematic review and meta-analysis. Lancet Psychiatry 4:295-301, 2017
20. Jongsma HE, Turner C, Kirkbride JB, et al.: International incidence of psychotic disorders, 2002-17: a systematic review and meta-analysis. The Lancet Public health 4:e229-e44, 2019
21. Firth J, Siddiqi N, Koyanagi A, et al.: The Lancet Psychiatry Commission: a blueprint for protecting physical health in people with mental illness. Lancet Psychiatry 6:675-712, 2019
22. Correll CU, Solmi M, Veronese N, et al.: Prevalence, incidence and mortality from cardiovascular disease in patients with pooled and specific severe mental illness: a large-scale meta-analysis of 3,211,768 patients and 113,383,368 controls. World Psychiatry 16:163-80, 2017
23. Vancampfort D, Stubbs B, Mitchell AJ, et al.: Risk of metabolic syndrome and its components in people with schizophrenia and related psychotic disorders, bipolar disorder and major depressive disorder: a systematic review and meta-analysis. World Psychiatry 14:339-47, 2015
24. Ferrari AJ, Stockings E, Khoo JP, et al.: The prevalence and burden of bipolar disorder: findings from the Global Burden of Disease Study 2013. Bipolar Disord 18:440-50, 2016
25. Moreira ALR, Van Meter A, Genzlinger J, et al.: Review and Meta-Analysis of Epidemiologic Studies of Adult Bipolar Disorder. J Clin Psychiatry 78:e1259-e69, 2017
26. Jääskeläinen E, Juola P, Hirvonen N, et al.: A systematic review and meta-analysis of recovery in schizophrenia. Schizophr Bull 39:1296-306, 2013
27. Leamy M, Bird V, Le Boutillier C, et al.: Conceptual framework for personal recovery in mental health: systematic review and narrative synthesis. Br J Psychiatry 199:445-52, 2011
28. Correll CU, Galling B, Pawar A, et al.: Comparison of Early Intervention Services vs Treatment as Usual for Early-Phase Psychosis: A Systematic Review, Meta-analysis, and Meta-regression. JAMA Psychiatry 75:555-65, 2018
29. Morrison AP, Law H, Barrowclough C, et al.: Programme Grants for Applied Research; in Psychological approaches to understanding and promoting recovery in psychosis and bipolar disorder: a mixed-methods approach: NIHR Journals Library. Southampton (UK) 2016
30. Meadows G, Brophy L, Shawyer F, et al.: REFOCUS-PULSAR recovery-oriented practice training in specialist mental health care: a stepped-wedge cluster randomised controlled trial. Lancet Psychiatry 6:103-14, 2019
31. Addington D, Anderson E, Kelly M, et al.: Canadian Practice Guidelines for Comprehensive Community Treatment for Schizophrenia and Schizophrenia Spectrum Disorders. Can J Psychiatry 62:662-72, 2017
32. Namjoshi MA, Buesching DP: A review of the health-related quality of life literature in bipolar disorder. Qual Life Res 10:105-15, 2001
33. Chen M, Fitzgerald HM, Madera JJ, et al.: Functional outcome assessment in bipolar disorder: A systematic literature review. Bipolar Disord 21:194-214, 2019
34. Porter ME, Larsson S, Lee TH: Standardizing Patient Outcomes Measurement. N Engl J Med 374:504-6, 2016
35. Obbarius A, van Maasakkers L, Baer L, et al.: Standardization of health outcomes assessment for depression and anxiety: recommendations from the ICHOM Depression and Anxiety Working Group. Qual Life Res 26:3211-25, 2017
36. Cumpston M, Li T, Page MJ, et al.: Updated guidance for trusted systematic reviews: a new edition of the Cochrane Handbook for Systematic Reviews of Interventions. Cochrane Database Syst Rev 10:Ed000142, 2019

37. Fitch K, Bochner A, Keller DS: Cost comparison of laparoscopic colectomy versus open colectomy in colon cancer. Current Medical Research and Opinion 33:1215-21, 2017
38. Terwee CB, Jansma EP, Riphagen, II, et al.: Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. Qual Life Res 18:1115-23, 2009
39. Reeve BB, Wyrwich KW, Wu AW, et al.: ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. Qual Life Res 22:1889-905, 2013
40. Nunnally JC: Psychometric Theory. New York: McGraw-Hill, 1978
41. Zerillo JA, Schouwenburg MG, van Bommel ACM, et al.: An International Collaborative Standardizing a Comprehensive Patient-Centered Outcomes Measurement Set for Colorectal Cancer. JAMA Oncol 3:686-94, 2017
42. Prinsen CAC, Mokkink LB, Bouter LM, et al.: COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res 27:1147-57, 2018
43. Faulkner S, Bee P: Experiences, perspectives and priorities of people with schizophrenia spectrum disorders regarding sleep disturbance and its treatment: a qualitative study. BMC Psychiatry 17:158, 2017
44. Connell J, Carlton J, Grundy A, et al.: The importance of content and face validity in instrument development: lessons learnt from service users when developing the Recovering Quality of Life measure (ReQoL). Qual Life Res 27:1893-902, 2018
45. Keetharuth AD, Brazier J, Connell J, et al.: Recovering Quality of Life (ReQoL): a new generic self-reported outcome measure for use with people experiencing mental health difficulties. Br J Psychiatry 212:42-9, 2018
46. Keetharuth AD, Taylor Buck E, Acquadro C, et al.: Integrating Qualitative and Quantitative Data in the Development of Outcome Measures: The Case of the Recovering Quality of Life (ReQoL) Measures in Mental Health Populations. Int J Environ Res Public Health 15, 2018
47. Conrad KJ, Yagelka JR, Matters MD, et al.: Reliability and Validity of a Modified Colorado Symptom Index in a National Homeless Sample. Mental Health Services Research 3:141-53, 2001
48. Park SG, Llerena K, McCarthy JM, et al.: Screening for negative symptoms: preliminary results from the self-report version of the Clinical Assessment Interview for Negative Symptoms. Schizophr Res 135:139-43, 2012
49. Mueser KT, Kim M, Addington J, et al.: Confirmatory factor analysis of the quality of life scale and new proposed factor structure for the quality of life scale-revised. Schizophr Res 181:117-23, 2017
50. Wiering B, de Boer D, Delnoij D: Patient involvement in the development of patient-reported outcome measures: a scoping review. Health Expect 20:11-23, 2017
51. Kwan B, Rickwood DJ: A systematic review of mental health outcome measures for young people aged 12 to 25 years. BMC Psychiatry 15:279, 2015
52. Jacobson NS, Truax P: Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 59:12-9, 1991
53. Jaeschke R, Singer J, Guyatt GH: Measurement of health status: Ascertaining the minimal clinically important difference. Controlled Clinical Trials 10:407-15, 1989
54. Gwaltney CJ, Shields AL, Shiffman S: Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: a meta-analytic review. Value Health 11:322-33, 2008
55. Kroenke K, Spitzer RL, Williams JB: The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 16:606-13, 2001
56. Conrad KJ, Yagelka JR, Matters MD, et al.: Reliability and validity of a modified Colorado Symptom Index in a national homeless sample. Ment Health Serv Res 3:141-53, 2001

57. Altman EG, Hedeker D, Peterson JL, et al.: The Altman Self-Rating Mania Scale. Biol Psychiatry 42:948-55, 1997
58. Buysse DJ, Yu L, Moul DE, et al.: Development and validation of patient-reported outcome measures for sleep disturbance and sleep-related impairments. Sleep 33:781-92, 2010
59. Ustun TB KN, Chatterji S.: Measuring health and disability: manual for WHO Disability Assessment Schedule (WHODAS 2.0). Geneva (CH): World Health Organization, 2010
60. Ravens-Sieberer U, Erhart M, Rajmil L, et al.: Reliability, construct and criterion validity of the KIDSCREEN-10 score: a short measure for children and adolescents' well-being and health-related quality of life. Qual Life Res 19:1487-500, 2010
61. Kroenke K, Spitzer RL, Williams JB: The PHQ-15: validity of a new measure for evaluating the severity of somatic symptoms. Psychosom Med 64:258-66, 2002
62. Waddell L, Taylor M: A new self-rating scale for detecting atypical or second-generation antipsychotic side effects. J Psychopharmacol 22:238-43, 2008
63. Beard C, Hsu KJ, Rifkin LS, et al.: Validation of the PHQ-9 in a psychiatric sample. J Affect Disord 193:267-73, 2016
64. Kroenke K, Spitzer RL, Williams JB, et al.: The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: a systematic review. Gen Hosp Psychiatry 32:345-59, 2010
65. Keetharuth A, Brazier, J., Connell, J., Carlton, J., Taylor Buck, E., Ricketts, T., Barkham, M.: Development and validation of the Recovering Quality of Life (ReQoL) outcome measures in Policy Research Unit in Economic Evaluation of Health and Care Interventions (EEPRU) Research Report 050. Edited by PRUiEEoHaC I: Universities of Sheffield and York, 2017
66. Sangha O, Stucki G, Liang MH, et al.: The Self-Administered Comorbidity Questionnaire: a new method to assess comorbidity for clinical and health services research. Arthritis Rheum 49:156-63, 2003

Figure 1. Summary of the development process for the psychotic disorders standard set

TABLE 1. Domain, outcome, definition, measure, timing of administration, and data sources

Domain	Outcome	Definition	Outcome Measure	Timing	Data Source
Symptoms					
	Depressive symptoms	Mood or emotional state that is marked by feelings such as depressed mood, hopelessness, worthlessness or guilt and a reduced ability to enjoy life	Patient Health Questionnaire 9-item (PHQ-9) (55)	Baseline and every six months	Patient
	Suicidal ideation and behavior	Suicidal ideation, suicidal thoughts or behaviors, suicide attempts, most often accompanied by intense feelings of hopelessness, depression, or self-destructive behaviors	Patient Health Questionnaire 9-item (PHQ-9) (55)	Baseline and every six months	Patient
	Positive symptoms	Change in thoughts or perceptions, including hallucinations, delusions, or disorganized thought, including anhedonia, blunting of emotions	Modified Colorado Symptom Index (MCSI) (56)	Baseline and every six months	Patient
	Negative symptoms*	A withdrawal or lack of function not expected in a healthy person, including blunting of affect, poverty of speech and thought, apathy, anhedonia, reduced social drive, loss of motivation, lack of social interest, and inattention to social or cognitive input	Recovering Quality of Life 20-item version (ReQoL-20) (45)	Baseline and every six months	Patient
	Mania/ Hypomania**	Abnormally elevated mood state characterized by symptoms such as inappropriate elation, increased irritability, severe insomnia, grandiose notions, increased speed and/or volume of speech, disconnected and racing thoughts, increased energy and activity level, and inappropriate social behavior	Altman Self-Rating Mania Scale (ASRM) (57)	Baseline and every six months	Patient
	Sleep quality	Problems with sleep resulting in decreased quality including difficulty falling asleep, difficulty staying asleep, early morning awakening, not feeling rested upon waking up	PROMIS Short Form V1.0 Sleep Disturbance 4a (PROMIS-Sleep) (58)	Baseline and every six months	Patient

	Relapse rate	Re-emergence of symptoms or disorder after partial or complete recovery	Hospitalizations (data point)	Baseline and every six months	Clinician/or patient
Recovery					
	Personal recovery	A personal unique process of changing one's attitudes, values, feelings, goals, skills and/or roles. A way of living a satisfying, hopeful and contributing life, according to CHIME Domains: Connectedness, Hope and Optimism, Identity, Meaning and Purpose and Empowerment	Recovering Quality of Life 20-item version (ReQoL-20) (45)	Baseline and every six months	Patient
	Quality of life	The individual's perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, including independence, experiencing a fuller range of emotions, and life satisfaction	Recovering Quality of Life 20-item version (ReQoL-20) (45)	Baseline and every six months	Patient
Functioning					
	Global functioning	An individual's social, occupational, and psychological functioning	WHODAS 2.0 (adult) (59) KIDSCREEN (adolescent) (60)	Baseline and every six months	Patient
	Social functioning	An individual's interactions with their environment, the quality of those interactions, and the individual's ability to fulfil their role within such environments as work, social activities and relationships with partners, families and/or friends.	WHODAS 2.0 (adult) (59) KIDSCREEN (adolescent) (60)	Baseline and every six months	Patient
	Role functioning	The ability of an individual to perform occupational activities, or a student's performance of functional tasks that support their participation in the academic aspects of school	WHODAS 2.0 (adult) (59) KIDSCREEN (adolescent) (60)	Baseline and every six months	Patient

	Physical health	Measure of physical health and well-being and overall satisfaction with physical health	Patient Health Questionnaire 15-item (PHQ-15) (61)	Baseline and every six months	Patient
Treatment					
	Side effects	Effect of medication prescribed, whether therapeutic or adverse, secondary to the one intended	Glasgow Antipsychotic Side-Effect Scale (GASS) (62)	Baseline and every six months	Patient

* specific to schizophrenia-spectrum disorders; ** specific to bipolar I disorder

TABLE 2. Psychometric properties of measures

Measure name	Complete name	# of items	Validity	Reliability	Sensitivity to Change	Reference
PHQ-9	Patient Health Questionnaire 9-item	9	++	+	**	Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med. Sep;16(9):606-13, 2001 (55) Beard, C, Hsu, KJ, Rifkin, LS, Busch, AB, & Bjorgvinsson, T. Validation of the PHQ-9 in a psychiatric sample. J Affect Disord. 193:267–273, 2016 (63) Kroenke K, Spitzer R, Williams J, and Lowe, B. The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: a systematic review. Gen Hosp Psychiatry 32:345-359, 2010 (64)
MCSI	Modified Colorado Symptom Index	14	+	++	×	Conrad KJ, Yagelka JR, Matters MD, Rich AR, Williams V, & Buchanan M. Reliability and validity of a modified Colorado Symptom Index in a national homeless sample. Ment Health Serv Res 3(3):141-153, 2001 (56)
ReQoL-20	Recovering Quality of Life 20-item version	20	NA Not a validated measure of negative symptoms Includes	++	*	Keetharuth AD, Brazier J, Connell J, Bjorner JB, Carlton J, Taylor Buck E, Ricketts T, McKendrick K,

			items that assess negative symptoms			Browne J, Croudace T, Barkham M. Recovering Quality of Life (ReQoL): a new generic self-reported outcome measure for use with people experiencing mental health difficulties. Br J Psychiatry. Jan;212(1):42-49, 2018(45) Keetharuth A, Brazier J, Connell J, Carlton J, Taylor Buck E, Ricketts T, & Barkham M. Development and validation of the Recovering Quality of Life (ReQoL) outcome measures. EEPRU Research Report 050. 2017. Edited by Interventions PRUiEEoHaC. Universities of Sheffield and York. http://scharr.dept.shef.ac.uk/eepru_wordpress/wp-content/uploads/2017/11/eepru-report-main-report-v3.pdf (65)
ASRM	Altman Self-Rating Mania Scale	5	NA	++	×	Altman EG, Hedeker D, Peterson JL, & Davis JM. The Altman Self-Rating Mania Scale. Biol Psychiatry, 42(10): 948–955, 1997 (57)
PROMIS-Sleep	PROMIS Short Form V1.0 Sleep Disturbance 4a	4	++	+	×	Buysse DJ, Yu L, Moul DE, Germain A, Stover A, Dodds NE, Johnston KL, Shablesky-Cade MA, Pilkonis PA. Development and

						validation of patient-reported outcome measures for sleep disturbance and sleep-related impairments. SLEEP 3 3(6):781-792, 2010. (58)
WHODAS 2.0 (adult functioning)	WHO Disability Assessment Schedule 2.0 – 12-item version	12	+	+	×	Ustun TB, Kostanjsek N, Chatterji S, et al. Measuring health and disability: manual for WHO Disability Assessment Schedule (WHODAS 2.0). Geneva (CH): World Health Organization 2010 (59)
KIDSCREEN-10 (adolescent functioning)		10	++	+	×	Ravens-Sieberer U, Erhart M, Rajmil L et al. Reliability, construct and criterion validity of the KIDSCREEN-10 score: a short measure for children and adolescents' well-being and health-related quality of life. Qual Life Res 19: 1487–1500, 2010 (60)
PHQ-15	Patient Health Questionnaire 15-item	15	+	++	*	Kroenke K, Spitzer RL, Williams JB. The PHQ-15: validity of a new measure for evaluating the severity of somatic symptoms. Psychosom Med, 64(2): 258-266, 2002(61) Kroenke K, Spitzer R, Williams J, and Lowe, B. The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: a systematic

						review. Gen Hosp Psychiatry 32:345-359, 2010 (64)
GASS	Glasgow Antipsychotic Side-Effect Scale	22	++	++	×	Waddell L, Taylor M. A new self-rating scale for detecting atypical or second-generation antipsychotic side effects. J Psychopharmacol 22(3):238-43, 2008 (62)

"++": Strong = acceptable interrater reliability ($r \geq 0.70$), internal consistency (Cronbach's $\alpha \geq 0.70$), and evidence of sensitivity to change across the majority of identified studies

"+": Mixed = only a single evaluation identified, or mixed evidence from several sources, or strong evidence only for certain items or sections

"-" Weak = majority of identified studies demonstrated below threshold evidence

"NA" = not assessed

"**": Well-established sensitivity to change

"*": Emerging evidence of sensitivity to change

"×": Further studies needed to assess sensitivity to change

TABLE 3. Risk adjustment factors

Patient Population	Measure	Supporting Information	Timing	Data Source
Demographic factors				
All patients	Year of birth	Not applicable (N/A)	Baseline	Patient-reported
All patients	Sex	Sex at birth	Baseline	Patient-reported
All patients	Gender identity	N/A	Baseline	Patient-reported
All patients	Sexual orientation	N/A	Baseline	Patient-reported
All patients	Socioeconomic status	**Adults** – highest level of education completed. **Adolescents** – proxy to be used: highest level of education completed by parents	Baseline; Transition to adult services; Annually if still in education	Patient-reported
All patients	Work / Education status	N/A	Baseline; Annually	Patient-reported
All patients	Housing status	N/A	Baseline; Annually	Patient-reported
All patients	Living arrangements	N/A	Baseline; Annually	Patient-reported
All patients	Ethnic minority/ Marginalization	N/A	Baseline	Patient-reported
Adult patients; Adolescent patients (where appropriate – see supporting information)	Contact with law enforcement	To be administered to adolescents only when appropriate to do so, and for whom this measure would not cause unnecessary distress. **Baseline** – ever been convicted (lifetime) **Annually** – ever been convicted (in the last 12 months)	Baseline; Annually	Patient-reported
Clinical factors				
All patients	Comorbidities	Based upon the Self-administered Comorbidity Questionnaire (66)	Baseline; Annually	Patient-reported
All patients	Hospitalizations	Number of lifetime hospitalizations related to the target condition.	Baseline	Administrative data
Adult patients; Adolescent patients (where appropriate – see supporting information)	Adverse life experiences	To be administered to adolescents only when appropriate to do so, and for whom this measure would not cause unnecessary distress. Tracked via the Adverse Childhood Experiences (ACE) Score. The scoring guide may be found at the same link.	Baseline; Transition to adult services	Patient-reported
Intervention factors				

All patients	Intervention setting	N/A		Baseline; Annually	Clinical
All patients	Intervention type	N/A		Baseline; Annually	Clinical