Patient-Reported Outcome Measures in Mental Health Clinical Research

A Descriptive Review in Comparison with Clinician-Rated Outcome Measures

Baandrup, Lone; Rasmussen, Jesper Østrup; Mainz, Jan; Videbech, Poul; Kristensen, Solvejg

Published in:
International Journal for Quality in Health Care

DOI (link to publication from Publisher):
10.1093/intqhc/mzab001

Publication date:
2022

Document Version
Accepted author manuscript, peer reviewed version

This is a pre-copyedited, author-produced version of an article accepted for publication in International Journal for Quality in Health Care following peer review. The version of record Lone Baandrup, Jesper Østrup Rasmussen, Jan Mainz, Poul Videbech, Solvejg Kristensen, Patient-reported outcome measures in mental health clinical research: a descriptive review in comparison with clinician-rated outcome measures, International Journal for Quality in Health Care, Volume 34, Issue Supplement_1, April 2022, Pages ii70–ii97, is available online at: https://doi.org/10.1093/intqhc/mzab001

REVIEW

Running title: PROMs in clinical research

Patient-reported outcome measures in mental health clinical research: a descriptive review in comparison with clinician-rated outcome measures

Lone Baandrup[1,2]

Jesper Østrup Rasmussen[3]

Jan Mainz[4,5,6]

Poul Videbech[7]

Solvejg Kristensen[4]

[1] Mental Health Center Copenhagen, Gentofte Hospitalsvej 15, 4., 2900 Hellerup, Denmark

[2] Department of Clinical Medicine, University of Copenhagen, Denmark

[3] Mental Health Center Ballerup, Copenhagen, Denmark

[4] Aalborg University Hospital, Psychiatry Aalborg, Denmark, DACS, Danish Center for Health Services Research, Aalborg University

[5] Clinical Institute, Aalborg University, Aalborg, Denmark

[6] Department for Community Mental Health, University of Haifa, Haifa, Israel

[7] Center for Neuropsychiatric Depression Research, Mental Health Centre Glostrup, Copenhagen, Denmark

Correspondence:

Lone Baandrup

Tuborgvej 235, DK-2400 København NV

Denmark

Tel +45 91165903

lone.baandrup@regionh.dk

Word count:

Abstract: 243

Text: 2959

Patient-reported outcome measures in mental health clinical research: a descriptive review in comparison with clinician-rated outcome measures

Running title: PROMs in clinical research

Abstract:

Purpose: To review how patient-reported outcomes measures in mental health clinical research complement traditional clinician-rated outcomes measures.

Data sources: Medline, Embase, PsycInfo and Scopus.

Study selection: Latest update of the literature search was conducted in August 2019, using a specified set of search terms to identify controlled and uncontrolled studies (published since 1996) of pharmacological or non-pharmacological interventions in adults (≥18 years) in hospital-based mental health care.

Data extraction: Two authors extracted data independently using a pre-designed extraction form.

Results of data synthesis: Among the 2962 publications identified, 257 were assessed by full text reading. A total of 24 studies reported in 26 publications were included in this descriptive review. We identified subjective and objective outcome measures, classified these according to the pharmacopsychometric triangle and compared them qualitatively in terms of incremental information added to the clinical study question. The data reviewed here from primarily depression and schizophrenia intervention studies show that results from patient-reported outcome measures and clinician-rated outcome measures generally point in the same direction. There was a relative lack of patient-reported outcome measures on functioning

and medication side effects compared with patient-reported outcome measures on symptom burden and health-related quality of life.

Conclusion: Patient-reported outcomes and clinician-rate outcomes supplement each other and at most times support identical study conclusions. Future studies would benefit from a more systematic approach towards use of patient-reported outcomes and a clearer rationale of how to weigh and report the results in comparison with clinician-rated outcomes.

Keywords: routine outcome measures, performance measures, clinical intervention, patient involvement, patient centered care

Introduction

It is estimated that each year 38% of the population in Europe suffers from a common mental disorder. Consequently, millions of people experience impairments in their everyday life and treatment is costly [1]. Mental health care services assume a central role in caring for patients with acute needs for psychiatric treatment and follow-up. Improving patient health status is the primary goal of healthcare, and patients, health care professionals, clinical managers, as well as health care planners and politicians take an interest in the quality of care provided - particularly in the health outcomes [2, 3].

Health outcomes concern all the effects of healthcare on individual patients or populations, where health can be regarded *"a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity*" [4]. Symptoms and symptom burden, adverse or side effects, as well as outcomes such as behavior, abilities and function, well-being and health-related quality of life are of importance to patients' overall status of health and affect daily life [5, 6]. The latter kind of outcomes goes beyond physiologic or laboratory measures and clinicians' observations and judgements, information which is *traditionally* reported by clinicians and used for outcome measurement. Such outcomes can be termed clinician-rated outcomes (CROs) and are typically used for clinical purposes as well as for performance measurement and quality of care improvement [7, 8]. It is known that discrepancies exist between patients' and clinicians' reports of symptoms and symptom burden as well as of functional status [9]. Consequently, the use of patient-reported outcomes (PROs) is of growing interest and use [10].

PROs are defined as *"any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else"* [11]. Accordingly, the PROs are free of observer bias, the patient is regarded an expert in the lived experience of his or her own health [12], and PROs must capture issues of importance to patients [13].

PROs are collected using patient reported outcome measures (PROMs). PROMs aim *"to capture patients' perspectives of health, illness and the effect of health care interventions in a reliable, valid acceptable and feasible way"* [14]. Hence, PROMs typically consist of multi-item scales for self-completion, where the

patient is asked to report on the extent to which certain pre-defined symptoms, side-effects, well-being issues or behavior have occurred during a certain time span. For example, if the patient's health problem impairs the ability to work [15], or to which extent the patient has felt active and vigorous during the past two weeks [16].

PROMs can either be disease-specific, condition-specific or generic. Disease-specific measures relate to a diagnostic group, e.g. surveying symptoms and symptom burden of patients diagnosed with schizophrenia or depression. Condition-specific measures relate to a specific condition (problem), e.g. sleep or cognition. Generic measures are designed for use with all patients unrelated to diagnosis or condition, e.g. functional status, quality of live or well-being [17]. The patient can fill in the PROMs at the hospital, in outpatient settings, or at home on paper or electronically [11]. The output mirrors the patient's perception of his or her health status at a single point in time, e.g., at diagnosis, before, during or after treatment, or during (long term) recovery. By surveying patients twice or more across time, the idea is to detect a change in health status attributable to an intervention [13, 18].

In measurement-based care (MBC) the scientific principle from controlled clinical trials are transferred to routine treatment to measure and improve the care of patients. Based upon data from patients with depression and anxiety, the pharmacopsychometric triangle has been established grouping PROMs into A) symptom and symptom burden, B) treatment side effects, and C) the resultant well-being and functioning. In MBC, it has been suggested that symptom burden and side effects are measured at all time points with addition of social functioning at 6 weeks and subjective well-being (quality of life) at 8 weeks [19]. Where the use of PROMs has shown their worth in large-scale MBC programs [20], little data exist regarding the comparison of traditional CROMs versus PROMs in research settings where the aim is not to compare MBC with traditional care (for review of this comparison, see reference [20]). The complex relation between use of patient-reported versus clinician-rated mental health outcomes have been the subject of previous reviews focusing on implementation of PROMs. Roe et al. suggested in a review of implementation and sustainability of PROMs several measures to enhance the efficiency of PROMs in adult mental health care

settings [21]. These suggestions included sufficient training, focus on administrative and logistic support, follow-up assessments and measures to reduce attrition rate. In another review the same author group concluded that implementation and sustainability of PROMs requires strong nationwide policy effort and support, otherwise implementation strategies are not systematic and consistent [22]. The level of agreement between self-reported and objective or provider-reported outcome measures has recently been evaluated in a cross-sectional study of 3,666 people with severe mental illness who participated in vocational rehabilitation programs [23]. Ratings of quality of life, functioning and illness management differed between groups and suggested differences in perspectives between consumers and providers regarding mental health outcomes.

When used in mental health clinical research, results from PROMs are most often reported in isolation and not directly compared with results from the applied CROMs. The US Food and Drug Administration (FDA) and the European Medicines Agency (EMA) have requested the use of PROMs in clinical trials for years [24], but still there is insufficient knowledge of the benefit or gain of using both PROMs and CROMs. It is not clear from the current guidelines how results from PROMs and CROMs, respectively, should be weighed against each other, and how they ought to be prioritized regarding choice of primary and secondary outcome measures. With this review we aim to investigate how results from PROMs and CROMS differ or agree in order to continuously develop and improve the use of PROMs in mental health clinical practice and research.

Study objectives

The objective of this review was to collect and characterize data on the comparison of PROMs and CROMs used in mental health care-based clinical research to improve knowledge of applicability of different PROMs and how they best supplement traditional CROMs. Based on these characteristics, we will formulate recommendations to guide future development within the field.

Methods

Design

The study is a systematic descriptive review.

Information sources, search strategy, study selection process and data extraction

Please refer to the online Supplementary for description.

Inclusion criteria

Studies were included in the review according to the following inclusion criteria: 1) randomized controlled trials and controlled/uncontrolled longitudinal studies of pharmacological and non-pharmacological interventions in hospital-based mental health care with use of CROMs and PROMs before and after the investigated intervention; 2) published between 1996 (inception of PubMed) and August 2019 in English or Scandinavian language; 3) primary diagnosis NOT abuse or dementia; 4) only studies on adults (≥18 years). The rationale behind inclusion criterium 1) was to focus not on how and which PROMs are implemented (this has been extensively reviewed elsewhere [21, 22]), but specifically on their role as outcome measures. We chose not to include studies examining patients with substance abuse or dementia since the main focus of the review was the two large illness domains schizophrenia and depression. We included only studies on adult populations since different PROMs and CROMs are applied in child and adolescent populations.

Exclusion criteria

The following study exclusion criteria were applied: 1) studies not carried out within hospital based (inpatient or outpatient) mental health care services; 2) editorials, commentaries, notes, clinical case reviews, opinion papers, and conference abstracts/posters/protocols/book chapters; 3) studies focusing on

patient reported experience or incidence measures; 4) studies describing PROMs and CROMs but not reporting any results.

Outcomes

We extracted data on the following items: 1) Aim of study; 2) study design including description of intervention; 3) study setting (inpatients or outpatients); 4) sample, age/gender, total N, funding and country; 5) timing of reporting; 6) CROM(s) (type, effect, quality); 7) PROM(s) (type, effect, quality); 8) effect of intervention according to CROMs and PROMs.

Appraisal of the quality of identified PROMs and CROMs

Since the purpose of a PROM is to illustrate the patient's experience and perspective, a PROM will not be a credible measure if there is no documentation of how it performs in the target population of patients. The quality of the identified PROMs and CROMs was assessed by two reviewers (LB and JR) using a pragmatic approach with the following ratings: ++) denoting that the PROM/CROM had previously been validated in a similar patient population; +) denoting that the PROM/CROM had previously been validated but not in a similar patient population; ?) insufficient information and -) PROM/CROM previously not systematically validated. The rating was based on the information given on each instrument in the respective papers.

Assessment of the quality of the included studies

Of the 24 included studies, 19 were randomized controlled trials. We assessed the quality of these using Cochrane's Risk of Bias Tool [25], and we assessed the quality of included non-randomized intervention studies using the tool ROBINS-I [26] (risk-of-bias plots were created using *robvis* [27]). All quality assessments were performed independently by two authors (LB and JØR) and discrepancies solved by consensus.

Narrative synthesis of study results

A narrative (descriptive) synthesis was carried out focusing on a qualitative analysis of the information obtained when adding PROMs to CROMs in clinical trials in mental health care.

Results

A total of 257 potentially relevant full text studies were identified for eligibility. We excluded 228 studies, the reasons for exclusion are listed in the on-line Supplementary Figure 1, with the most frequent one (n=91) being studies where the design was not in agreement with the inclusion criteria of the current review. Consequently, a total of 24 studies reported in 26 publications were included in the current review. Results are summarized in Tables 1-3 reporting, respectively, characteristics of included studies and summaries of CROMs and PROMs. Eight of the included studies were conducted in Europe [28-35], six studies were multinational trials [36-41], four studies from the US [42-45], three studies from Asia [46-48], two studies from Canada [49, 50] and one study from Australia [51]. Studies were mainly concerned with major depressive disorder (11 studies) [29, 33, 38, 39, 41, 43, 44, 46, 47, 49, 50] or schizophrenia (11 studies) [28, 30-32, 34-37, 40, 42, 45]. One study reported on obsessive-compulsive disorder (OCD) [48] and one on severe mental illness [51]. Fourteen of the studies had <200 subjects included [28, 31-36, 44, 45, 47-51], eight studies ≥200 subjects [29, 30, 37, 39-42, 46] and 2 studies had ≥1000 subjects included [38, 43]. The majority of the studies investigated a pharmacological intervention (15 studies) [28, 29, 34-44, 49, 50], eight studies investigated various psychological interventions [30-33, 45-47, 51], and one study repetitive transcranial magnetic stimulation [48]. All studies included outpatients except for a single study [38] including participants from inpatient settings also. In the majority of the studies, the time frame was well above the suggested 8 weeks (according to the pharmacopsychometric triangle [19]) for measuring well-being/health-related quality of life. Only three studies [36, 44, 48] had follow-up times shorter than 8 weeks, but the outcomes measured at these shorter follow-up intervals were related to symptom burden (6 weeks) and side effects (4 weeks) and thus still compatible with the suggested follow-up intervals [19].

In Table 4-5 the studies have been listed according to the principles of the pharmacopsychometric triangle [19] to provide an overview of the applied PROMs: whether they measure subjective symptom burden, subjective side effects of the intervention, restoration of social functioning, or well-being (health-related quality of life). Based on the data in Table 4-5, we were able to adopt five comparisons: 1) symptom burden PROM versus symptom burden CROM; 2) symptom burden CROM versus health-related quality of life PROM; 3) symptom burden CROM versus side effects PROM; 4) symptom burden CROM versus social functioning PROM, and 5) social functioning PROM versus social functioning CROM:

1) For the patient-reported versus clinician-rated burden of symptoms, we found 11 studies that included this comparison. Only three [32, 43, 45] out of these 11 studies reported discrepancy between the CROM and the PROM. The three studies examined, respectively, an educational intervention in schizophrenia [32], different antidepressants in treatment-resistant depression [43], and cognitive training in schizophrenia [45].

2) For the clinician-rated burden of symptoms versus patient-reported quality of life we found 13 studies reporting this comparison out of which only one showed discrepancy. This was a medication trial in schizophrenia reporting improvement in clinician-rated severity of cognitive dysfunction which was not replicated in health-related quality of life [28].

3) For the clinician-rated burden of symptoms versus patient-reported side effects, we found six studies [28, 32, 34, 43, 45, 46] reporting this comparison. Two of these showed discrepancies: pharmacist-based shared decision making interventions in depression [46] and a medication intervention in schizophrenia [42].

4) The clinician-rated burden of symptoms versus patient-reported social functioning we found seven studies [31, 32, 37, 39, 44, 49, 50] reporting on this comparison, none of these with discrepancy between findings.

5) For the patient-reported social functioning versus clinician-rated social functioning we found one study [37] reporting this association but no discrepancy between measures.

The PROMs applied were all well-established tools, however, in six studies [29-31, 46, 50, 51] none of the applied PROMs had previously been validated in the specific patient population in question (Table 2-3). Eighteen [28-33, 36-41, 43-48] out of the 24 included studies were randomized trials, five [34, 35, 42, 49, 50] studies were open-label and one [51] had a naturalistic design. Most studies were associated with low or unclear risk of bias in the assessed domains (see Supplementary Figure 2). The non-randomized studies were associated with the highest risk of bias with some studies showing serious risk of bias (see Supplementary Figure 3).

Discussion

In this narrative review of 24 studies investigating experimental treatment interventions in mental health care, we report that PROMs and CROMs most often point in the same direction when extracting only the quantitative parts of the data as typically summarized in some form of sum score. This supports the view that PROMs and CROMs supplement each other examining different aspects of outcome measures, i.e. symptom burden, adverse effects, health-related quality of life, even though the results from clinical trials are not markedly changed when using both PROMs and CROMs as compared to only one of them. Using PROMs both in the clinic and in research most be expected to support patient involvement, self-management and the relation between patient and clinician.

We did not identify any specific pattern characterizing the studies that reported different results in PROMs versus CROMs. The high level of agreement between clinician and patient ratings is in line with a recent study investigating this question in a sample of patients with treatment-resistant depression finding moderate-strong relationship between the assessment tools [52]. Likewise, in that study, few predictors of discordance between CROMs and PROMs were identified though chronicity was associated with greater agreement. An important perspective is that some of the potentially beneficial aspects of using PROMs are not adequality addressed in the included studies of this review, i.e. how the use of PROMs influences the

development of a therapeutic rapport and adherence to treatment. In a pilot study of a newly developed Danish PROM battery [53] for use in mental disorders that the author group performed recently, it was a consistent finding that the patients considered the PROMs very useful to improve their interaction with the team of health care professionals and their perceived outcome of the treatment effort [54].

The results of this review indicate a relative lack of use of PROMs on functioning and medication side effects compared with PROMs on symptom burden and health-related quality of life. Data on side effects as CROMs were not extracted for this review. Only one of the included studies examined both clinician-rated and patient-reported level of social functioning whereas several studies examined patient-reported social functioning alone, indicating a choice towards patient-reported tools for this outcome domain.

The research community ought to request more focus on including social functioning CROMs and PROMs to get a broader view on this difficult-to-measure domain that has such important impact on both prognosis and well-being/quality of life. This is consistent with recovery models that emphasize functional rather than symptomatic improvement [55]. Social functioning is a complex construct and scales often do not distinguish adequately between functioning and psychopathology. Considering this, we encourage future studies to include an improved understanding and a better definition of how remission or recovery might be reflected in the applied CROMs and PROMs. Concepts of remission and recovery are important in order to assure that reductions on a given rating scale (CROM or PROM) are relevant from a patient perspective and reflected in increased levels of functioning. Connecting CROMs and different aspects of recovery has recently been attempted by Best et al. [56] analyzing a data set of 971 subjects with schizophrenia using baseline ratings from four studies. They reported that various symptom domains were differentially associated with personal versus functional recovery, e.g. affective symptoms were markedly more associated with personal recovery than with objective functioning, and thus separated functional and personal recovery as distinct domains. Our data set was too heterogenous to meta-analyze and thus we cannot confirm a pattern like this. But what we can confirm is a need to standardize how CROMs and PROMs and ideally a combination of these measures can be used to identify patient-relevant outcomes as

remission and recovery. A recommendation for future attempts to move the field forward would clearly be to focus on how and when combinatorial CROMs and PROMs could be applied. The nature of the current data set clearly demonstrates that CROMs and PROMs are widely reported as distinct and very separate measures with no clear connection.

The nature of the relation between applied CROMs and PROMs was heterogenous across included studies namely PROMs being used to measure symptom burden, medication side effects, heath-related quality of life and social functioning. We did not in this review include studies which were MBC-based only, i.e. applying only PROMs and no CROMs. It was not very clear from the included studies exactly what was the purpose of including a certain PROM and how the results were prioritized in comparison to CROMs.

A large randomized study that evaluated outcome measures using both CROMs and PROMs is represented by the STAR*D trial [57]. This was the largest study ever on treatment of depression and examined sequential steps of pharmacotherapy. The results of this study have been published in several papers reporting on different steps and different subpopulations from the study. Unfortunately, we were not able to include STAR*D in the current review, because we could not identify results from both CROMs and PROMs reported from the same subset of the sample. However, Ishak et al. [58] summarized the results and reported that despite a significant impact on quality of life, functioning, and depressive symptom severity, a substantial proportion of participants still suffered from reduced patient-reported quality of life and functional impairment after treatment, which was particularly evident for non-remitters. This is consistent with the findings from Dunlop and colleagues [43] reporting a poorer level of agreement between CROM and PROM for response than for remission in treatment-resistant depression.

Strengths of this review include the systematic literature search and systematic methods of study selection and data extraction. To our knowledge, this is the first review of its kind. It adds to the knowledge base of what to expect when including both PROMs and CROMs in clinical research in mental disorders regarding supplementary or complementary results from the two types of instruments.

Limitations of the current study: CROMs and PROMs from the same study are frequently reported in different publications, not always cross-referenced, which made it difficult to obtain the corresponding CROM to a specific PROM. For this reason, we had to exclude articles where a CROM was either not included in the study or not available for extraction. Results were not always reported for each outcome measure but only as correlations or associations between reported outcomes which made it impossible to perform quantitative or numerical comparisons. Some CROMs require training or certification before use and the extent of this may vary between studies introducing some uncertainty about the precision of scoring. Generally, the psychometric validity among the variety of used rating scales might not have been sufficiently validated even though many of the applied tools were claimed to have been validated previously. We did not examine in further details what the procedures of validation for each questionnaire were comprised of. The inclusion of both randomized and non-randomized studies implied heterogeneity between included studies which could potentially affect the conclusions of the study. However, when considering the comparison of direction of outcome, which was the main focus of this review, there was no obvious disagreement with the general pattern when looking at randomized and non-randomized studies separately.

According to the original concept of including PROMs in clinical trials, it has been highlighted in the Food and Drug Administration guidance on patient reported outcomes for labeling and promotional claims that it is necessary to ensure responsiveness of applied PROs [11]. This should be done by demonstrating that the PRO scores are sensitive to actual changes in clinical or health status and by determining the minimal important difference to assist in interpreting statistically significant PRO results in clinical trials [59]. It has also been emphasized that the minimal important difference has to be established for a particular study population [59]. We did not evaluate the responsiveness to change of the PROMs applied in the included studies in the current review, but merely looked at whether the instrument had previously been validated in the patient population in question.

We conclude from this review that PROMs and CROMs supplement each other in mental health care-based clinical research and at most times point in the same direction. Sometimes, there is additional quantitative informative results to gain when using both PROMs and CROMs to measure efficiency in clinical trials. A gain that is inherent despite similar quantitative results is the different perspectives and points of view represented by PROMs and CROMs, respectively. Future trials need to include PRO measures of social functioning and medication side effects in addition to PRO measures of symptom burden and health-related quality of life. Consequently, these measures are also likely to be informative for inclusion in mental health clinical practice. In addition, future work needs to focus on how PROMs and CROMs can be better integrated to define outcomes measures of remission and recovery.

Acknowledgments

None.

Conflicts of interest

The authors report no conflicts of interest.

Funding

This work was supported by the Obel Family Foundation under grant No 28020. Furthermore, the Danish Clinical Registries; Aalborg University; and Aalborg University Hospital, Psychiatry are acknowledged for financial support.

Author contributions

LB and SK designed the review, PV and JM contributed to conception and design. LB, JØR and SK performed the literature search and selection of studies. LB and JØR extracted the data and performed the quality

assessments. LB, JØR, SK, PV and JM all contributed to analysis and interpretation of data. LB drafted the manuscript and all other authors revised it critically for important intellectual content. All authors approved the final version of the manuscript and the decision to submit. All authors agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work were appropriately investigated and resolved.

References

1. Wittchen HU, Jacobi F, Rehm J, Gustavsson A, Svensson M, Jonsson B, et al. The size and burden of mental disorders and other disorders of the brain in Europe 2010. Eur Neuropsychopharmacol 2011;21:655-79.

2. OECD Health Ministers. The next generation of health reforms: ministerial statement. Avaliable from www.oecd.org/health/ministerial/ministerial-statement-2017.pdf. [Last accessed 18.07.2019]. Organisation for Economic Co-operation and Development (OECD). 2017:1-17.

3. Browne JP, Cano SJ, Smith S. Using Patient-reported Outcome Measures to Improve Health Care: Time for a New Approach. Med Care 2017;55:901-4.

4. World Health O. World Health Organization - Mental health: a state of well-being. Avaliable from http://www.who.int/features/factfiles/mental_health/en/ [Last accessed 18.07.2019]. World Health OrganizationGeneva, Schwitzerland. 2014.

5. Campbell SM, Shield T, Rogers A, Gask L. How do stakeholder groups vary in a Delphi technique about primary mental health care and what factors influence their ratings? Qual Saf Health Care 2004;13:428-34.

6. Jorgensen M, Mainz J, Svendsen ML, Nordentoft M, Voldsgaard I, Baandrup L, et al. Improving quality of care among patients hospitalised with schizophrenia: a nationwide initiative. B J Psych Open 2015;1:48-53.

7. Baandrup L, Cerqueira C, Haller L, Korshoj L, Voldsgaard I, Nordentoft M. The Danish Schizophrenia Registry. Clin Epidemiol 2016;8:691-5.

8. Videbech P, Deleuran A. The Danish Depression Database. Clin Epidemiol 2016;8:475-8.

9. Registries for Evaluating Patient Outcomes, 3rd edition A User's Guide. Gliklich RE, Dreyer NA, Dreyer MB, editors. Rockville (MD), USA: Agency for Healthcare Research and Quality (US); 2014, p. 2-208.

10. Basch E, Bennett A, Pietanza MC. Use of patient-reported outcomes to improve the predictive accuracy of clinician-reported adverse events. J Natl Cancer Inst 2011;103:1808-10.

11. Administration FaD. Draft Guidance for Industryon Patient-reported Outcome measures: Use in Medicinal Product Development to Support Labeling Claims. US Department of Health and Human Services 2009.

12. Kendrick T, El-Gohary M, Stuart B, Gilbody S, Churchill R, Aiken L, et al. Routine use of patient reported outcome measures (PROMs) for improving treatment of common mental health disorders in adults. Cochrane Database Syst Rev. 2016;7:CD011119.

13. Marshall S, Haywood K, Fitzpatrick R. Impact of patient-reported outcome measures on routine practice: a structured review. J Eval Clin Pract 2006;12:559-68.

14. Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. Health Technol Assess 1998;2:i-74.

15. Mundt JC, Marks IM, Shear MK, Greist JH. The Work and Social Adjustment Scale: a simple measure of impairment in functioning. Br J Psychiatry 2002;180:461-4.

16. Topp CW, Ostergaard SD, Sondergaard S, Bech P. The WHO-5 Well-Being Index: a systematic review of the literature. Psychother Psychosom 2015;84:167-76.

17. Dorflinger LH, Knudsen JL. [An introduction to patient reported outcome measures in quality assessment]. Ugeskr Laeger 2015;177(19).

18. Murphy M, Salisbury C, Hollinghurst S. Can the outcome of primary care be measured by a Patient Reported Outcome Measure? Br J Gen Pract 2014;64:647-8.

19. Bech P, Timmerby N. An overview of which health domains to consider and when to apply them in measurement-based care for depression and anxiety disorders. Nord J Psychiatry 2018:1-7.

20. Fortney JC, Unutzer J, Wrenn G, Pyne JM, Smith GR, Schoenbaum M, et al. A Tipping Point for Measurement-Based Care. Psychiatr Serv 2017;68:179-188.

21. Gelkopf M, Mazor Y, Roe D. A systematic review of patient-reported outcome measurement (PROM) and provider assessment in mental health: goals, implementation, setting, measurement characteristics and barriers. Int J Qual Health Care 2020; Mar 11;mzz133.

22. Roe D, Mazor Y, Gelkopf M. Patient-reported outcome measurements (PROMs) and provider assessment in mental health: a systematic review of the context of implementation. Int J Qual Health Care. 2019; 2019 Nov 14;mzz084.

23. Gal G, Shadmi E, Hoter-Ishay G, Gelkopf M, Roe D. Comparing outcome measures of persons with severe mental illness in vocational rehabilitation programs: a dual perspective of consumers and providers. Int J Qual Health Care 2020;Mar 31;mzaa030.

24. Administration FaD. Draft Guidance for Industryon Patient-reported Outcome measures: Use in Medicinal Product Development to Support Labeling Claims. US Department of Health and Human Services 2009.

25. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011;343:d5928.

26. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ 2016;355:i4919.

27. McGuinness LA, Higgins JPT. Risk-of-bias VISualization (robvis): An R package and Shiny web app for visualizing risk-of-bias assessments. Res Synth Methods. 2020; April 2020.

28. Baandrup L, Fagerlund B, Glenthoj B. Neurocognitive performance, subjective well-being, and psychosocial functioning after benzodiazepine withdrawal in patients with schizophrenia or bipolar disorder: a randomized clinical trial of add-on melatonin versus placebo. Eur Arch Psychiatry Clin Neurosci 017;267:163-71.

29. Fantino B, Moore N. The self-reported Montgomery-Åsberg depression rating scale is a useful evaluative tool in major depressive disorder. BMC Psychiatry 2009;9:26

30. Holt RI, Hind D, Gossage-Worrall R, Bradburn MJ, Saxon D, McCrone P, et al. Structured lifestyle education to support weight loss for people with schizophrenia, schizoaffective disorder and first episode psychosis: the STEPWISE RCT. Health Technol Assess 2018;22:1-160.

31. Magliano L, Fiorillo A, Malangone C, De Rosa C, Maj M. Patient functioning and family burden in a controlled, real-world trial of family psychoeducation for schizophrenia. Psychiatr Serv 2006;57:1784-91.

32. Merinder LB, Viuff AG, Laugesen HD, Clemmensen K, Misfelt S, Espensen B. Patient and relative education in community psychiatry: A randomized controlled trial regarding its effectiveness. Soc Psychiatry Psychiatr Epidemiol 1999;34:287-94.

33. Meuldijk D, Carlier IV, Van Vliet IM, van VT, Wolterbeek R, van Hemert AM, et al. The clinical effectiveness of concise cognitive behavioral therapy with or without pharmacotherapy for depressive and anxiety disorders; a pragmatic randomized controlled equivalence trial in clinical practice. Contemp Clin Trials 2016;47:131-8.

34. Pietrini F, D'Anna G, Tatini L, Talamba GA, Andrisano C, Calderani E, et al. Changes in attitude towards LAI antipsychotic maintenance treatment: A two-year follow-up study. Eur Pschiatry 2018;53:58-65.

35. Pietrini F, Spadafora M, Talamba GA, Godini L, Lelli L, Arcabasso S, et al. The effects of switching from oral to LAI antipsychotic treatment on subjective experience of schizophrenic and schizoaffective patients: Preliminary results. Int J Psychiatry Clin Pract 2015;19(2):106-13.

36. Canuso CM, Grinspan A, Kalali A, Damaraju CV, Merriman U, Alphs L, et al. Medication satisfaction in schizophrenia: A blinded-initiation study of paliperidone extended release in patients suboptimally responsive to risperidone. Int Clin Psychopharmacol 2010;25:155-64.

37. Dunayevich E, Buchanan RW, Chen CY, Yang J, Nilsen J, Dietrich JM, et al. Efficacy and safety of the glycine transporter type-1 inhibitor AMG 747 for the treatment of negative symptoms associated with schizophrenia. Schizophr Res 2017;182:90-7.

38. Florea I, Danchenko N, Brignone M, Loft H, Rive B, Abetz-Webb L. The effect of vortioxetine on health-related quality of life in patients with major depressive disorder. Clin Ther 2015;37:2309-23.

39. Francois C, Nielsen R, Danchenko N, Williams V, Lancon C. The impact of antidepressant treatments on family functioning in adults with major depressive disorder: a post hoc comparison of vortioxetine and agomelatine. Curr Med Res Opin 2017;33:1057-66.

40. Kane JM, Sanchez R, Baker RA, Eramo A, Peters-Strickland T, Perry PP, et al. Patient-Centered Outcomes with Aripiprazole Once-Monthly for Maintenance Treatment in Patients with Schizophrenia: Results From Two Multicenter, Randomized, Double-Blind Studies. Clin Schizophr Relat Psychoses 2015;9:79-87.

41. Locklear JC, Sveds+ter H, Datto C, Endicott J. Effects of once-daily extended release quetiapine fumarate (quetiapine XR) on quality of life and sleep in elderly patients with major depressive disorder. J Affect D 2013;149:189-95.

42. Andorn A, Graham J, Csernansky J, Newcomer JW, Shinde S, Muma G, et al. Monthly Extended-Release Risperidone (RBP-7000) in the Treatment of Schizophrenia: Results From the Phase 3 Program. J Clin Psychopharmacol 2019;39:428-433.

43. Dunlop BW, McCabe B, Eudicone JM, Sheehan JJ, Baker RA. How well do clinicians and patients agree on depression treatment outcomes? Implications for personalized medicine. Human Psychopharmacology 2014;29:528-36.

44. Mathew SJ, Gueorguieva R, Brandt C, Fava M, Sanacora G. A randomized, double-blind, placebo-controlled, sequential parallel comparison design trial of adjunctive riluzole for treatment-resistant major depressive disorder. Neuropsychopharmacology 2017;42:2567-74.

45. Treichler EBH, Thomas ML, Bismark AW, Hochberger WC, Tarasenko M, Nungaray J, et al. Divergence of subjective and performance-based cognitive gains following cognitive training in schizophrenia. Schizophr Res 2019;210:215-20.

46. Aljumah K, Hassali MA. Impact of pharmacist intervention on adherence and measurable patient outcomes among depressed patients: A randomised controlled study. BMC Psychiatry 2015;15:2019.

47. Guo T, Xiang YT, Xiao L, Hu CQ, Chiu HF, Ungvari GS, et al. Measurement-Based Care Versus Standard Care for Major Depression: A Randomized Controlled Trial With Blind Raters. The American journal of psychiatry 2015;172:1004-13.

48. Haghighi M, Shayganfard M, Jahangard L, Ahmadpanah M, Bajoghli H, Pirdehghan A, et al. Repetitive Transcranial Magnetic Stimulation (rTMS) improves symptoms and reduces clinical illness in patients suffering from OCD--Results from a single-blind, randomized clinical trial with sham cross-over condition. . J Psychiatr Res 2015;68:238-44.

49. Cao B, Park C, Subramaniapillai M, Lee Y, Iacobucci M, Mansur RB, et al. The efficacy of vortioxetine on anhedonia in patients with major depressive disorder. Front Psychiatry 2019;10:17.

50. Sarfati D, Evans VC, Tam EM, Woo C, Iverson GL, Yatham LN, et al. The impact of fatigue and energy on work functioning and impairment in patients with major depressive disorder treated with desvenlafaxine. International Clinical Psychopharmacology 2017;32:343-9.

51. Meehan T, Robertson S. Impact of the Mental Health Nurse Incentive Programme on patient functioning. Int J Ment Health Nurs 2015;24(1):75-81.

52. Hershenberg R, McDonald WM, Crowell A, Riva-Posse P, Craighead WE, Mayberg HS, et al. Concordance between clinician-rated and patient reported outcome measures of depressive symptoms in treatment resistant depression. J Affect Disord 2020;266:22-29.

53. Kristensen S, Mainz J, Baandrup L, Bonde M, Videbech P, Holmskov J, et al. Conceptualizing patient-reported outcome measures for use within two Danish psychiatric clinical registries: description of an iterative co-creation process between patients and healthcare professionals. Nord J Psychiatry. 2018;72(6):409-19.

54. Kristensen S. Evaluering af PRO-Psykiatri. Basereret på erfaringer fra Psykiatrien i Region Nordjylland. In: Nordjylland PiR, editor. https://psykiatrirndk//media/Hospitaler/Psykiatrien/Brugerinddragelse/Evalueringsrapport-PRO-Psykiatriashx?la=da2020 [last accessed 21.04.2020].

55. Leamy M, Bird V, Le Boutillier C, Williams J, Slade M. Conceptual framework for personal recovery in mental health: systematic review and narrative synthesis. Br J Psychiatry. 2011;199:445-52.

56. Best MW, Law H, Pyle M, Morrison AP. Relationships between psychiatric symptoms, functioning and personal recovery in psychosis. Schizophr Res 2020 Jun 30;S0920-9964(20)30378-9.

57. Sinyor M, Schaffer A, Levitt A. The sequenced treatment alternatives to relieve depression (STAR*D) trial: a review. Can J Psychiatry. 2010;55:126-35.

58. IsHak WW, Mirocha J, Pi S, Tobia G, Becker B, Peselow ED, et al. Patient-reported outcomes before and after treatment of major depressive disorder. Dialogues Clin Neurosci 2014;16:171-183.

59. Revicki DA, Cella D, Hays RD, Sloan JA, Lenderking WR, Aaronson NK. Responsiveness and minimal important differences for patient reported outcomes. Health Qual Life Outcomes. 2006;4:70.

Table 1 Summary of included studies

Study	Aim	Design	Setting	Sample	Total N	Study period	Fu
Aljumah et al. 2015	To assess whether pharmacist-based interventions based on shared decision making (SDM) improve adherence and patient-related outcomes	Randomized controlled trial	Outpatients	Major depressive disorder Age: Not reported Male: 100 (45%)	239	3 months	N re

Andorn et al. 2019 + Dhanda et al. 2019	To describe the long-term impact of RBP-7000 on HRQoL, subjective well-being, treatment satisfaction and medication preference	Multi-centre Phase III single-arm open-label study	Outpatients	Schizophrenia Age: 45.1 (SD not reported) Male: 326 (68%)	482	52 weeks	Th sp su pl in
Baandrup et al. 2017	To examine how melatonin and benzodiazepine withdrawal affect cognition, subjective well-being, and psychosocial functioning	Randomized, double-blind trial	Outpatients	Schizophrenia and bipolar disorder Age: 47.4 (SD 8.6) vs. 49.0 (SD 12.1) Male: 45 (56%)	80	24 weeks	Fu in gr
Canuso et al. 2010	To assess antipsychotic medication satisfaction in patients who were switched from risperidone to paliperidone ER	Randomized controlled trial	Outpatients	Schizophrenia insufficiently treated with risperidone Age: 40.6 (SD 11.9) Male: 111 (56%)	191	6 weeks	Th su M Sc
Cao et al. 2019	To examine the efficacy of vortioxetine on anhedonia in major	Post-hoc analysis of open-label	Outpatients	Major depressive disorder Age: 38.9 (SD 12.9) Male: 33 (35%)	95	8 weeks	Fu A/

	depressive disorder	study					
Dunayevich et al. 2017	To determine the safety and efficacy of AMG 747, an oral inhibitor of glycine transporter type-1 (GlyT1), as an add-on to antipsychotic therapy	Randomized controlled trial	Outpatients	Schizophrenia with predominant negative symptoms Age: 43.9 (SD 10.5) Male: 156 (67%)	232	12 weeks	A
Dunlop et al. 2014	To evaluate the effect of self-reported/clinician-rated agreement on patient-level outcomes	Pooled data from three randomized, double-blind, placebo-controlled trials, data from phase B: Open-label treatment with one of five antidepressants, flexibly dosed, along with the co-	Outpatients	Treatment- resistant depression Age: 44.0 (SD 11.0) vs. 44.8 (SD 11.0) Male: 659 (32%)	2075	8 weeks	B a P

		administration of a single-blind placebo					
Fantino et al. 2009	To assess the psychometric properties of the 9-item, patient-administered version of the Montgomery-Åsberg Depression Rating Scale (MADRS-S)	Data came from a multicenter, double blind, randomized clinical trial	Outpatients	Major depressive disorder Age: 39.5 (SD 12) Male: 92 (33%)	278	8 weeks	H pr th ur
Florea et al. 2015	To describe the effect of vortioxetine on HRQoL in MDD patients by using patient-reported outcomes instruments	5 short-term (6-8 weeks), randomized studies	Both outpatient and inpatient settings	Major depressive disorder (MDD) Mean age ranged between 42 and 47 years in the individual studies Male: 34%	2155 (vortioxetine) vs. 1316 (placebo)	6-8 weeks	Th ar ar sp Lu Ta Co
Francois et al. 2017	To examine the impact of vortioxetine and agomelatine on family functioning	Randomized, double-blind trial	Both inpatients and outpatients	Major depressive disorder with inadequate response to antidepressant Age: 46.3 (SD 12.0)	376	12 weeks	Fu A,

				Male: 95 (25%)			
Guo et al. 2015	To compare measurement-based care with standard treatment in major depression	Randomized controlled trial, with assessors blind to protocol and treatment group	Outpatients	Non-psychotic major depression Age: 41.1 (SD 12.1) Male: 43 (36%)	120	24 weeks	Su go ar gr
Haghighi et al. 2015	To examine if repetitive Transcranial Magnetic Stimulation (rTMS) improves symptoms and reduces illness severity in patients suffering from treatment-resistant	Randomized, single-blind, sham, controlled clinical trial with cross-over	Outpatients	Treatment-resistant OCD Age: 35.9 (SD 11.0) Male: 12 (57%)	21	4 weeks	N

	OCD						
Holt et al. 2018	To develop a lifestyle intervention and to evaluate its clinical effectiveness, cost-effectiveness, delivery and acceptability	Two-arm, analyst-blind, parallel-group, randomized controlled trial	Outpatients	First episode psychosis, schizophrenia or schizoaffective disorder Age: 40.0 (SD 11.3) vs 40.1 (SD 11.5) Male: 210 (51%)	414	12 months	Th fu In Re Te As pr
Kane et al. 2015 + Fleischhacker 2014	To assess the efficacy, safety, and tolerability of aripiprazole once-monthly (400 mg) for the maintenance treatment compared with oral aripiprazole	Double-blind, active-controlled, non-inferiority study	Outpatients	Schizophrenia according to DSM-IV-TR Age: 41.7 (SD 10.4) Male: 160 (60%) vs. 168 (63%)	662 responders to oral aripiprazole were randomized	38 weeks	O Lu
Locklear et al. 2013	To investigate the effects of once-daily extended-release (XR) quetiapine	Multicenter, double-blind,	Outpatients	Major depressive disorder Age: 71.2 (SD 4.9) vs. 71.3 (SD 4.6)	338	9 weeks	As

		parallel-group, placebo-controlled trial		Male: 51 (30%) vs. 49 (30%)			
Magliano et al. 2006	To explore impact of psychoeducational intervention on patients' personal and social functioning	Randomized controlled trial	Outpatients	Families of consumers with schizophrenia Age: 36.9 (SD 8.2) vs. 34.1 (SD 7.8) Male: 29 (69%) vs. 24 (83%)	71 families	6 months	M ar In
Mathew et al. 2017	To examine the efficacy and safety of riluzole vs. placebo as an adjunct to antidepressant medication	Randomized, double-blind controlled trial	Outpatients	Patients with major depressive disorder with an inadequate response to antidepressant medication Age: 43.3 (SD 12.7) vs. 47.3 (SD 12.1) vs. 44.5 (SD 12.2) Male: 50 (46%)	104	Two phases of 4 weeks	Th su N M H
Meehan et al. 2015	To improve levels of care, reduce the likelihood of unnecessary admissions, assist in keeping people with severe illnesses feeling well	Naturalistic study	Outpatients	Severe mental illness Age 43.0 (SD 13.5) Male: 50 (60%)	84	Ranged from a minimum of 3 weeks to a maximum of 28 weeks (median 15 weeks)	Th G
Merinder	To evaluate the	Randomized	Outpatients	Schizophrenia	46	12 months	H

et al. 1999	effectiveness of an educational intervention for patients with schizophrenia and their relatives	controlled trial		Age: 35.9 Male: 24 (52%)			D
Meuldjjk et al. 2016	Comparison between concise forms of CBT and/or pharmacotherapy	Pragmatic randomized controlled equivalence trial	Outpatients	Mild to moderate anxiety and/or depression Age 36.5 (SD 12.3) Male: 71 (39%)	182	12 months	P
Pietrini et al. 2015	To evaluate effects of switching from SGA oral to LAI formulation	Prospective, longitudinal, open-label, non-randomized, single-arm, observational study	Outpatients	Schizophrenia or schizoaffective disorder Age 40.7 (SD 10.9) Male: 14 (54%) Participants were about to be switched from oral to the equivalent maintenance regimen with LAI	27	6 months	N
Pietrini et al. 2018	To present real-world evidence on the effects of switching from oral to long-acting injectable (LAI) antipsychotic maintenance treatment (AMT)	Prospective, longitudinal, open-label, nonrandomized, single-arm, observational study	Outpatients	Adult patients with schizophrenia in need of long-term antipsychotic treatment Age: 38.4 (SD 11.1) Male: 26 (61%)	50	24 months	N au co

Sarfati et al. 2017	To examine the impact of the symptoms fatigue and low energy on work functioning in patients with major depressive disorder	Secondary analysis of open-label study	Outpatients	Major depressive disorder Age: 39.2 (SD 10.9) Male: 15 (43%)	35	8 weeks	Fu fr
Treichler et al. 2019	1) To determine the effect of cognitive training on subjective cognitive difficulties and cognitive performance	Randomized controlled trial	Not reported	Schizophrenia or schizoaffective disorder Age: 34.5 (SD 12.1) vs 35.7 (SD 13.0) Male: 22 (47%)	46	10 weeks	G re

Table 2 Summary of intervention and clinician-rated outcome measures (CROMs) in included studies

Study	Intervention	CROM	Type	Effect	Quality [a]	CROM	Type
Aljumah et al. 2015	Enhancing patients' involvement in decision making by assessing their beliefs and knowledge about antidepressants; 2 visits (baseline and 3 months follow-up). Versus control group (usual pharmacy services)	MADRS (Montgomery Åsberg Depression Rating Scale)	Disease-specific	No difference between intervention groups	++		
Andorn et al. 2019 + Dhanda et al. 2019	RBP-7000 (120 mg) - a once-monthly subcutaneous extended-release risperidone formulation	PANSS (Positive and Negative Syndrome Scale) total score, Positive Scale, Negative Scale and General Psychopathology Scale scores	Disease-specific	Over 12 months of exposure, mean PANSS scores continued to improve in rollover participants and remained stable among de novo participants	++	CGI (Clinical Global Impression)	Generic

Baandrup et al. 2017	Prolonged-release melatonin 2 mg versus placebo	BACS (Brief Assessment of Cognition in Schizophrenia)	Disease-specific	BACS composite and subscale scores (except motor speed) significantly improved in parallel with benzodiazepine dose reduction, but there was no additional effect of melatonin	++	PSP (Personal and Social Performance Scale)	Generic
Canuso et al. 2010	Immediate or delayed initiation of paliperidone XR	PANSS (Positive and Negative Syndrome Scale) total score	Disease-specific	For both groups, mean total PANSS scores improved from baseline to week 2 and at all subsequent time points, but no significant differences between groups were observed at any time point	++	CGI (Clinical Global Impression)	Generic
Cao et al.	Vortioxetine (10–	SNAITH	Disease-specific	Significant	++	MDRS	Disease-

2019	20mg, flexibly dosed) daily for 8 weeks	Hamilton pleasure scale		baseline to endpoint improvement in anhedonia factor score ($p < 0.0001$)		(Montgomery Åsberg Depression Rating Scale) anhedonia factor	specific
Dunayevich et al. 2017	Adults diagnosed with schizophrenia stabilized on antipsychotic medication randomized (2:2:2:3) to orally receive daily AMG 747 (5mg, 15mg, or 40mg) or placebo	PANSS (Positive and Negative Syndrome Scale) Negative Symptom Factor Score	Disease-specific	At week 12, the mean decrease from baseline in PANSS NSFS was significantly greater with 15-mg AMG whereas the 5-mg and 40 mg groups did not show statistically significant difference from placebo	++	PSP (Personal and Social Performance Scale)	Condition-specific
Dunlop et al. 2014	Open-label treatment with one of five antidepressants versus single-blind placebo	HAM-D17 (Hamilton Depression Rating Scale)	Disease-specific	Not reported in isolation	++		
Fantino et al.	Comparing escitalopram with	MADRS (Montgomery	Disease-specific	Not reported in isolation	++	CGI-S (Clinical	Generic

2009	citalopram	Åsberg Depression Rating Scale)				Global Impression of Severity)	
Florea et al. 2015	Vortioxetine (5, 10, 15 and 20 mg/d) versus placebo	MADRS (Montgomery Åsberg Depression Rating Scale)	Disease-specific	Not reported in isolation	++		
Francois et al. 2017	Eligible patients at baseline were directly switched from their previous treatment by randomization (1:1) to vortioxetine (10–20 mg/day) or agomelatine (25–50 mg/day) for 12 weeks of double-blind treatment	CGI-S (Clinical Global Impression of Severity)	Generic	Vortioxetine significantly superior to agomelatine	++	MDRS (Montgomery Åsberg Depression Rating Scale)	Disease-specific
Guo et al. 2015	Measurement-based care (guideline- and rating scale based decisions), or standard treatment (clinicians' choice decisions).	HAM-D17 (Hamilton Depression Rating Scale)	Disease-specific	Both the response rate and the remission rate were significantly higher in the intervention	++	YMRS (Young Mania Rating Scale)	Disease-specific

	Pharmacotherapy was restricted to paroxetine (20–60mg/day) or mirtazapine (15–45mg/day) in both groups.			group			
Haghighi et al. 2015	In addition to standardized SSRI- or clomipramine medication at therapeutic dosages and CBT, all patients were treated with rTMS for two weeks and with an rTMS sham condition for two weeks.	CGI (Clinical Global Impression Scale)	Generic	CGI severity values decreased significantly over time. The significant Time by Group Interaction showed that CGI severity values decreased over time under the rTMS condition, but not under the sham-condition	+		
Holt et al. 2018	Intervention group: (1) four 2.5-hour group-based structured lifestyle self-management	BPRS (Brief Psychiatric Rating Scale)	Condition-specific	No significant difference	++		

	education sessions, 1 week apart; (2) multimodal fortnightly support contacts; (3) three 2.5-hour group booster sessions at 3-monthly intervals, post core sessions. Control group: usual care						
Kane et al. 2015 + Fleischhacker 2014	Aripiprazole once-monthly 400 mg versus oral aripiprazole (10–30 mg/day)	Relapse	Generic	Kaplan–Meier estimated impending relapse rates at week 26 were 7.12% for AOM and 7.76% for oral ARI. This excluded the predefined non-inferiority margin of 11.5%	-	PANSS (Positive and Negative Syndrome Scale)	Disease-specific
Locklear et al. 2013	Quetiapine XR (flexible dosing 50-300 mg/day) versus placebo	MADRS (Montgomery Åsberg Depression Rating Scale)	Disease-specific	Total score reduced (improved) in intervention group	++		

Mathew et al. 2017	Patients were randomly assigned to adjunctive treatment with riluzole (50 mg twice per day) or placebo	MADRS (Montgomery Åsberg Depression Rating Scale)	Disease-specific	Treatment groups did not differ in mean change in MADRS scores	++	CGI-S (Clinical Global Impression of Severity)	Generic
Meehan et al. 2015	Across the study period, patients received between one and 44 sessions (mean = 8.5, SD 18.2)	HoNOS (Health of Nation Outcome Scales), higher is worse	Generic	The mean total HoNOS score decreased (improved)	++	GAF (Global Assessment of Functioning)	Generic
Merinder et al. 1999	8-session psychoeducational program for patients with schizophrenia and their relatives versus usual treatment in outpatient psychiatric clinics	BPRS (Brief Psychiatric Rating Scale)	Disease-specific	Trend improvement (0.07) in the intervention group	++	GAF (Global Assessment of Functioning)	Generic
Meuldijk et al. 2016	7-session concise version of CBT and/or pharmacotherapy versus longer standard care	CGI (Clinical Global Impression)	Generic	Reduced in both groups	++		
Pietrieni	Change from oral	MADRS	Disease-specific	Improvement of	++	PANSS	Disease-

et al. 2015	second-generation antipsychotic to long-acting injectable (n=18: olanzapine; n=8 paliperidone)	(Montgomery Åsberg Depression Rating Scale)		several psychometric indexes		(Positive and Negative Syndrome Scale)	specific
Pietrini et al. 2018	At the time of enrolment (T0), all patients were under a stabilized therapy with a single oral second-generation antipsychotic (SGA) and were switched to the equivalent maintenance regimen with the long-acting formulation of the same antipsychotic	PANSS (Positive and Negative Syndrome Scale)	Disease-specific	Significant improvement after one year of LAI antipsychotic maintenance therapy, with stable results after two years	++	MADRS (Montgomery Åsberg Depression Rating Scale)	Disease-specific
Sarfati et al. 2017	Patients were treated for 8 weeks with desvenlafaxine starting at 50 mg per day. Dose could be increased to 100 mg per day at week 2 or later at the discretion of the clinic psychiatrist	MADRS (Montgomery Åsberg Depression Rating Scale)	Disease-specific	Significant improvements	++		

Treichler et al. 2019	30h of auditory-targeted cognitive training	MCCB (MATRICS Consensus Cognitive Battery)	Disease-specific	Not reported	++	SANS/SAPS (Scale for Assessment of Negative/Positive Symptoms)	Disease-specific

Table 3 Summary of patient-reported outcome measures (PROMs) in included studies

Study	PROM	Type	Effect	Quality [a]	PROM	Type
Aljumah et al. 2015	EQ-5D (Health-related QoL)	Generic	Estimated weights for EQ-5D showed no significant differences between groups	+	TSQM (Treatment Satisfaction Questionnaire for Medication)	Condition-specific
Andorn et al. 2019 + Dhanda et al. 2019	EuroQoL 5D 5-Level (EQ-5D-5L)	Generic	EQ-5D-5L index remained stable from baseline to 52 weeks follow-up	++	MSQ (Medication Satisfaction Questionnaire)	Condition-specific
Baandrup et al. 2017	WHO-5 (WHO-5 well-being index)	Generic	Neither benzodiazepine withdrawal nor treatment group affected subjective well-being	+	SWN (Subjective Well-being on Neuroleptics)	Disease-specific
Cao et al.	WHO-5 (WHO-5	Generic	Improvements in the	++	SDS (Sheehan Disability	Generic

2019	well-being index)		SHAPS and the MADRS anhedonia factor correlated with improvements in quality of life (i.e., WHO-5) ($p < 0.0001$)		Scale)	
Canuso et al. 2010	MSQ (Medication Satisfaction Questionnaire)	Condition-specific	Paliperidone ER was associated with a significant increase (improvement) in MSQ scores from baseline; there were no statistically significant between-group differences at this and other time points	++	SF-36 SF-36 (The Short Form-36 Health Survey) Mental health composite score	Generic
Dunayevich et al. 2017	Q-LES-Q-18 (Quality of Life Enjoyment and Satisfaction Questionnaire)	Generic	Changes from baseline to week 12 in the Q-LES-Q-18 showed evidence of greater efficacy of 15-mg AMG 747 compared with placebo (p=0.058)	++	SDS (Sheehan Disability Scale work/school item)	Condition-specific

Dunlop et al. 2014	IDS-SR (Depressive Symptomatology Self-rated)	Disease-specific	Not reported in isolation	++		
Fantino et al. 2009	MADRS (Montgomery Åsberg Depression Rating Scale)-S = patient administered version	Disease-specific	Not reported in isolation, focusing on psychometric properties	+		
Florea et al. 2015	SF-36 (The Short Form-36 Health Survey)	Generic	Treatment with vortioxetine was associated with significant clinically meaningful improvements in HRQoL, including specific improvements on the SF-36 mental health domains of vitality social	++	Q-LES-Q-18 (Quality of Life Enjoyment and Satisfaction Questionnaire)	Generic

			functioning, role emotional and mental health			
Francois et al. 2017	DFFS (Depression and Family Functioning Scale)	Disease-specific	Vortioxetine was superior to agomelatine by 2.5 points at week 12 ($p<.05$)	++	SDS (Sheehan Disability Scale)	Generic
Guo et al. 2015	QIDS-SR (Quick Inventory of Depressive Symptomatology–Self-Report)	Disease-specific	Only reported in intervention group – decrease from baseline to follow-up	++	The Frequency, Intensity, and Burden of Side Effects Rating scale	Condition-specific
Haghighi et al. 2015	Y-BOCS (Yale-Brown Obsessive-Compulsive Scale); self-rating	Disease-specific	Y-BOCS values decreased significantly over time. No group differences were observed. The significant Time by Group Interaction showed that Y-BOCS values decreased over time in the rTMS condition, but not in the sham-	++		

			condition			
Holt et al. 2018	EuroQoL 5D 5-Level (EQ-5D-5L)	Generic	Greater improvement among control participants at 12 months, with a difference of 4.4 points (p = 0.028)	+	SF-26 (The Short Form questionnaire-36 items), from which eight domains of quality of life (QoL) were derived	Generic
Kane et al. 2015 + Fleischhacker 2014	DAI (The Drug Attitude Inventory)	Condition-specific	Mean DAI remained stable across all treatment phases, no difference between treatment arms	++	MAQ (The Medication Adherence Questionnaire)	Condition-specific
Locklear et al. 2013	Q-LES-Q-SF (Quality of Life and Satisfaction Questionnaire Short Form)	Generic	Improvement in intervention group	++	PSQI (Pittsburgh Sleep Quality Index)	Condition-specific
Magliano et al. 2006	FPQ (Family Problem Questionnaire)	Condition-specific	The average level of family burden improved in both groups	+	SNQ (Social Network Questionnaire)	Condition-specific
Mathew et al. 2017	IDS-SR (Inventory of Depressive	Disease-specific	Treatment groups did not differ	++	CPFQ (Cognitive and Physical Functioning Questionnaire)	Condition-specific

	Symptomatology-Self Rated)					
Meehan al. 2015	DASS-21 (Depression, Anxiety and Stress Scale-21)	Disease-specific	Scores for all three DASS-21 subscales decreased (improved) from baseline to follow-up	+		
Merinder et al. 1999	IS (Insight Scale)	Condition-specific	No differences between groups	++	VSSS (Verona Service Satisfaction Scale)	Condition-specific
Meuldijkj et al. 2016	Brief Symptom Inventory, based on the SCL-90 (Symptom Check-List 90)	Generic	Improved in both groups	++	SF-36 (The Short Form-36 Health Survey)	Generic
Pietrieni et al. 2015	SWN-K (Subjective well-being on neuroleptics, short version)	Disease-specific	Widespread improvement in all of the five SWN-K subscales	++	SF-36 (The Short Form-36 Health Survey)	Generic
Pietrini et al. 2018	SWN-K (Subjective Well-Being Under	Condition-specific	Significant improvement after one year of LAI antipsychotic	++	DAI-10 (The Drug Attitude Inventory short version)	Condition-specific

	Neuroleptics scale short form)		maintenance therapy, with stable results after two years			
Sarfati et al. 2017	PROMIS (Patient-Reported Outcomes Measurement Information System) Fatigue scale	Condition-specific	Significant improvement in Montgomery–Åsberg Depression Rating Scale scores as well as in fatigue measures following treatment	+	SDS (Sheehan disability scale)	Condition-specific
Treichler et al. 2019	MIC-SR (Measure of Insight into Cognition-Self Report)	Disease-specific	MIC-SR did not significantly change over time, and there was no significant change associated with TCT participation	++	PHQ-9 (Patient Health Questionnaire, 9 item depression sub scale)	Condition-specific

[a] Quality: ++) the PROM has been systematically validated in a similar patient population; +) validated but in a different patient population; ?) insufficient information; -) not systematically validated

Table 4 Identification of the comparisons within the pharmacopsychometric triangle

Study	Nature of comparison	Qualitative comparison CROM versus P
Aljumah et al. 2015	Symptom burden CROM vs. HRQoL PROM Symptom burden CROM vs. side effects PROM	For symptom severity (CROM) and HRQoL (PROM) there were no between intervention groups. After 6 months, intervention group patients showed statistically 18% in adherence to antidepressants and 6% in treatment satisfa concern beliefs and general beliefs about medicines.
Andorn et al. 2019 + Dhanda et al. 2019	Symptom burden CROM vs. HRQoL PROM Symptom burden CROM vs.side effects PROM	Stable measures of both PANSS, CGI and HRQoL. Improvement in was not visible from the CROMs used.
Baandrup et al. 2017	Symptom burden CROM vs. HRQoL PROM Social functioning CROM vs. HRQoL PROM	Improvement in CROM with benzodiazepine dose reduction, no d
Cao et al. 2019	Symptom burden CROM vs. HRQoL PROM Symptom burden CROM vs. social functioning PROM	Vortioxetine improved measures of anhedonia, which significantly correlated with improvements in function.
Canuso et al. 2010	Symptom burden CROM vs. HRQoL PROM Symptom burden CROM vs. side effects PROM	Improvement in all scales from baseline to follow-up, no differen
Dunayevich et al. 2014	Symptom burden CROM vs. HRQoL PROM	The efficacy (CROM) of the AMG 747 15mg dose was supported b particular the Q-LES-Q18 total score and the Q-LES-Q-18 social do change from baseline in the social domain subscale was correlate the change from baseline in the PANSS NSFS.
Dunlop et al. 2014	Symptom burden CROM vs. symptom burden PROM	In this sample of patients with TR MDD, levels of agreement were and PROM for the definitions of response to monotherapy AD tre definition of remission.
Fantino et al.	Symptom burden CROM vs. symptom burden	The correlation between MADRS-S (PROM) and physicians' MADR

2009	PROM	0.54, p < 0.001).
Florea et al. 2015	Symptom burden CROM vs. HRQoL PROM	Treatment with vortioxetine was associated with significant clinic improvements in HRQoL supporting the efficacy profile in depres vortioxetine has demonstrated a statistically significant differenc
Francois et al. 2017	Symptom burden CROM vs. symptom burden PROM Symptom burden CROM vs. social functioning PROM	The better DFFS, the better outcome on the other scales.
Guo et al. 2015	Symptom burden CROM vs. symptom burden PROM Symptom burden CROM vs. side effects PROM	Improvement in both HAM-D and QIRS-SR in intervention group
Haghighi et al. 2015	Symptom burden CROM vs. symptom burden PROM	Changes in symptoms were apparent in both the self-ratings and improvement over time and a significant time by group interactio magnetic stimulation.
Holt et al. 2018	Symptom burden CROM vs. social functioning PROM	Better family functioning was associated with better functional s symptoms.
Kane et al. 2015 + Fleischhacker 2014	Symptom burden CROM vs. side effects PROM	Results equal across groups for PANNS (CROM) and PROMs
Locklear et al. 2013	Symptom burden CROM vs. symptom burden PROM Symptom burden CROM vs. HRQoL PROM	Improvement in depressive symptoms (CROM) and PROMs inclu
Magliano et al.	Symptom burden CROM vs. social functioning PROM	In the intervention group improvement in global level of disabilit in getting a job, social interests and management of conflicts (CR

		practice of support (PROM).
Mathew et al. 2017	Symptom burden CROM vs. HRQoL PROMs	No significance in either objective outcomes or PROM.
Mehaan et al. 2015	Symptom burden CROM vs. symptom burden PROM	Highly significant improvements on all of the domains assessed
Merinder et al. 1999	Symptom burden CROM vs. symptom burden PROM Symptom burden CROM vs. social functioning PROM	The intervention group showed a trend improvement in symptor improvement in satisfaction with relatives' involvement (PROM) symptom burden PROM.
Meuldijk et al. 2016	Symptom burden CROM vs. symptom burden PROM Symptom burden CROM vs. HRQoL PROM	Significant improvements in all outcome measures in both interv between groups.
Pietrieni et al.	Symptom burden CROM vs. HRQoL PROM	Significant improvements in psychometric indexes (CROM) and p of treatment (PROM) in both initial remitters and non-remitters.
Pietrini et al. 2018	Symptom burden CROM vs. HRQoL PROM and side effects PROM	Significant improvement in both objective outcomes and PROMs
Sarfati et al. 2017	Symptom burden CROM vs. symptom burden PROM and social functioning PROM	Fatigue measures were significantly associated with improvement in some work functioning measures.
Treichler et al. 2019	Symptom burden CROM vs. symptom burden PROM and social functioning PROM	Subjective cognitive difficulties did not significantly improve follo among participants who showed improvements in cognitive perf relationships among measures of subjective cognitive difficulties and objective cognitive performance were detected.

CROM: Clinician-rated outcome measure; PROM: Patient-reported outcome measure; HRQoL: Health-related quality of life

Table 5 Summary of the CROM/PROM comparisons according to the pharmacopsychometric triangle

Study	Symptom burden PROM	Symptom burden CROM	Side effects PROM	Well-being (HR QoL) PROM	Social functioni PROM
Aljumah et al. 2015		MADRS (Montgomery Åsberg Depression Rating Scale)	TSQM (Treatment Satisfaction Questionnaire for Medication) MMAS (The Morisky Medication Adherence Scale)	EQ-5D (EuroQoL 5D 5-Level)	
Andorn et al. 2019 + Dhanda et al. 2019		PANSS (Positive and Negative Syndrome Scale) total score CGI (Clinical Global Impression Scale)	MSQ (Medication Satisfaction Questionnaire) POM (Preference of Medication Questionnaire)	EQ-5D-5L (EuroQoL 5D 5-Level) SF-36v2 (Short-Form 36-item Questionnaire, Version 2 SWN-S (Subjective Well-being Under Neuroleptic Treatment-Short Version)	
Baandrup et al. 2017		BACS (Brief Assessment of Cognition in Schizophrenia)		SWN-S (Subjective Well-being Under Neuroleptic Treatment-Short Version) WHO-5 (WHO-5 well-being scale)	
Cao et al.		SNAITH Hamilton		WHO-5 (WHO-5 well-	SDS (Sheehan disa

2019		pleasure scale MDRS (Montgomery Åsberg Depression Rating Scale) anhedonia factor		being scale)	scale)
Canuso et al. 2010		PANSS (Positive and Negative Syndrome Scale) total score	MSQ (Medication Satisfaction Questionnaire)	SF-36 (The Short Form-36 Health Survey) composite score	
Dunayevich et al. 2017		PANSS (Positive and Negative Syndrome Scale) Negative Symptom Factor Score		Q-LES-Q-18 Quality of Life Index (Quality of Life Enjoyment and Satisfaction Questionnaire)	SDS (Sheehan Dis Scale work/schoo item)
Dunlop et al. 2014	IDS-SR (Depressive Symptomatology Self-rated)	HAM-D17 (Hamilton Depression Rating Scale)			
Fantino et al. 2009	MADRS (Montgomery Åsberg Depression Rating Scale)-S = patient administered	MADRS (Montgomery Åsberg Depression Rating Scale)			

	version	CGI-S (Clinical Global Impression of Severity)			
Florea et al. 2015		MADRS (Montgomery Åsberg Depression Rating Scale)		SF-36 (The Short Form-36 Health Survey) Q-LES-Q (Quality of Life Enjoyment and Satisfaction Questionnaire) EQ-5D (EuroQoL 5 dimension)	
Francois et al. 2017		CGI-S (Clinical Global Impression of Severity) MADRS (Montgomery Åsberg Depression Rating Scale)			DFFS (Depression Family Functionin Scale) SDS (Sheehan disa scale)
Guo et al. 2015	QIDS-SR (Quick Inventory of Depressive Symptomatology Self-report)	HAM-D (Hamilton Depression Rating Scale)	The Frequency, Intensity, and Burden of Side Effects Rating scale		
Haghighi et al. 2015	Y-BOCS (Yale-Brown Obsessive-	CGI (Clinical Global Impression			

	Compulsive Scale) - self-rating	Scale)			
Holt et al. 2018		BPRS (Brief Psychiatric Rating Scale)		EuroQol-5 (EuroQoL 5D 5-Level) SF-36 (The Short Form-36 Health Survey)	
Kane et al. 2015 + Fleischhacker 2014		Relapse PANSS (Positive and Negative Syndrome Scale)	DAI (Drug Attitude Questionnaire) MAQ (Medication Adherence Questionnaire)		
Locklear et al. 2013	PSQI (Pittsburgh Sleep Quality Index)	MADRS (Montgomery Åsberg Depression Rating Scale)		Q-LES-Q-SF (Quality of Life and Satisfaction Questionnaire Short Form)	
Magliano et al. 2006		BPRS (Brief Psychiatric Rating Scale) Assessment of disability (derived from the Disability Assessment Schedule)			FPQ (Family Probl Questionnaire) SNQ (Social Netw Questionnaire)
Mathew et al. 2017	IDS-SR (Inventory of Depressive	CGI-S (Clinical Global Impression			CPFQ (Cognitive a Physical Function

	Symptomatology-Self Rated)	Scale) MADRS (Montgomery Åsberg Depression Rating Scale)			Questionnaire)
Meehan et al. 2015	DASS-21 (Depression, Anxiety and Stress Scale-21)	HoNOS (Health of Nation Outcome Scales)			
Merinder et al. 1999	IS (Insight Scale)	BPRS (Brief Psychiatric Rating Scale)			VSSS (Verona Serv Satisfaction Scale
Meuldjjk et al. 2016	Brief Symptom Inventory based on the SCL-90	CGI (Clinical Global Impression)		SF-36 (The Short Form-36 Health Survey)	
Pietrieni et al. 2015		MADRS (Montgomery Åsberg Depression Rating Scale) PANSS (Positive and Negative Syndrome Scale)		SWN-K (Subjective well-being on neuroleptics, short version) SF-36 (The Short Form-36 Health Survey)	
Pietrini et al. 2018		PANSS (Positive and Negative Syndrome Scale) MADRS	DAI (Drug Attitude Questionnaire)	SWN-K (Subjective Well-Being Under Neuroleptics scale short form)	

		(Montgomery Åsberg Depression Rating Scale)			
Sarfati et al. 2017	PROMIS (Patient-Reported Outcomes Measurement Information System) Fatigue scale	MADRS (Montgomery Åsberg Depression Rating Scale)			SDS (Sheehan disa scale)
Treichler et al. 2019	MIC-SR (Measure of Insight into Cognition-Self Report) PHQ-9 (Patient Health Questionnaire, 9 item depression sub scale)	MCCB (MATRICS Consensus Cognitive Battery) SANS/SAPS (Scale for Assessment of Negative/Positive Symptoms)			

CROM: Clinician-rated outcome measure; PROM: Patient-reported outcome measure; HRQoL: Health-related quality of life