

A Systematic Review of Mental Health Measurement Scales for Evaluating the Effects of Mental Health Prevention Interventions

Josefien JF Breedvelt (MSc)^{1 (a1)}, Victoria Zamperoni (MSc)^{2 (a1)}, Emily South (MSc)³, Eleonora Uphoff (PhD)⁴, Simon Gilbody (PhD)⁵, Claudi LH Bockting (PhD)⁶, Rachel Churchill (PhD)⁴, & Antonis A Kousoulis (MD)²

a1: Note these authors are joint first authors

Corresponding Author:

Victoria Zamperoni

Mental Health Foundation

Colechurch House, 1 London Bridge Walk

London

SE1 2SX

vzamperoni@mentalhealth.org.uk

02078031113

Author Affiliations

1. Mental Health Foundation, London, UK and Amsterdam UMC, Department of Psychiatry, University of Amsterdam, Amsterdam, Netherlands
2. Mental Health Foundation, London, UK
3. Centre for Reviews and Dissemination, University of York, York, UK
4. Cochrane Common Mental Disorders, University of York, York, UK
5. Mental Health and Addictions Research Group, Department of Health Sciences, University of York, UK
6. Amsterdam UMC, Department of Psychiatry, Institute for Advanced Study, University of Amsterdam, Amsterdam, Netherlands

Abstract

Background: Consistent and appropriate measurement is needed in order to improve understanding and evaluation of preventative interventions. This review aims to identify individual-level measurement tools used to evaluate mental health prevention interventions to inform harmonisation of outcome measurement in this area.

Methods: Searches were conducted in PubMed, PsychInfo, CINAHL, Cochrane and OpenGrey for studies published between 2008 and 2018. that aimed to evaluate prevention interventions for common mental health problems in adults and used at least one measurement scale (PROSPERO CRD42018095519). For each study, mental health measurement tools were identified and reviewed for reliability, validity, ease-of-use, and cultural sensitivity.

Results: 127 studies were identified that used 65 mental health measurement tools. Most were used by a single study (57%, N = 37) and measured depression (N = 20) or overall mental health (N = 18). The most commonly used questionnaire (15%) was the Centre for Epidemiological Studies Depression Scale (CES-D). A further 125 tools were identified which measured non- mental health specific outcomes.

Conclusions: There was little agreement in measurement tools used across mental health prevention studies, which may hinder comparison across studies. Future research on measurement properties and acceptability of measurements in applied and scientific settings could be explored. Further work on supporting researchers to decide on appropriate outcome measurement for prevention would be beneficial for the field.

Key Words: prevention; mental health; outcome assessment; review

Introduction

Depression, the most prevalent diagnosis of mental health problems, currently represents one of the leading causes of disability globally,¹ and the number of people living with a common mental health problem (defined by the National Institute of Health and Care Excellence as depression and anxiety disorders) is increasing.²⁻⁴ Prevention of common mental health problems, such as depression, remain an important area of focus for improving public health and reducing demand on public services.

Taking a prevention-based approach to mental health problems involves a public-health model of action, with interventions directed toward entire populations (universal), groups known to have higher risk of mental health problems (selective), and those already displaying early symptoms (indicated).⁵ Mental health prevention interventions are therefore interventions which aim to prevent common mental health problems at one of those three levels.

Both prevention and promotion interventions consider a wide range of risk and protective factors which can be understood at the individual, family, community, and population-wide levels. Characteristics may also include age, gender, culture, income, experiences, and education.⁶ Currently, a wide range of measures appear to be used by academics and stakeholders with little guidance on key, or core outcomes to be used by researchers. In order to improve comparability of findings in mental health prevention research, it has been recommended that data collection and evaluation methods are aligned and more consistent use is made of robust standardised measures to allow for the improved use of data across larger and smaller-scale projects, routine outcome measurement, and trials.⁷

A comprehensive overview of the range of measurement tools currently used in the literature to evaluate mental health prevention interventions can serve as a valuable resource for researchers, practitioners, and policymakers and represents an initial step toward more standardised measurement across prevention projects. Utilising a consistent suite of robust measurement tools can enable more direct comparison of results across studies, allowing for easier consolidation of research findings and aiding the ability to draw generalisable conclusions that inform policy and practice.

This review aims to systematically identify the range of measurement tools currently used to evaluate interventions which aim to prevent common mental health problems.

Methods

Search Strategy

The protocol for the systematic review was registered on PROSPERO (CRD42018095519) in May 2018. Recent studies which aimed to evaluate preventative interventions for common mental health problems (depression, anxiety disorders, PTSD, and OCD).⁴ were identified through a series of electronic searches conducted in PubMed, PsychINFO, CINAHL, Cochrane, and OpenGrey databases. Search terms were a combination of terms in the title or abstract indicative of: adults, non-clinical populations, mental health, prevention, and measurement. Search strategies are presented in Appendix I. Appropriate standardised vocabulary and truncation were used for each database. Only those studies published in English, between January 2008 and May 2018 were included.

Inclusion and Exclusion Criteria

To be eligible for inclusion, studies had to: be published between January 2008-May 2018 (last search date: 18/05/2018), involve evaluation (using at least one measurement scale) of an intervention with a non-clinical adult population (mean age of 16-65 without a mental health diagnosis) and have a primary aim of improving mental health and preventing a common mental health problem. This included universal, selective, and indicated mental health prevention intervention.⁵

Studies were excluded from analysis if: they had a primary aim of preventing conditions or behaviours other than common mental health problems, including psychosis, schizophrenia, bipolar disorder, eating disorders, dementia, substance abuse, gambling, suicide, or domestic abuse; they had a primary aim other than the prevention of mental health problems (e.g. promotion of subjective wellbeing or overall health-related quality of life); they worked with a population that met clinical threshold for a mental health problem at baseline; they did not involve a direct intervention (e.g. prevalence studies, measure development and validation studies); they did not report use of a measurement scale.

Within included studies, specific measurement tools were excluded if they: focussed on populations secondary to the intervention (e.g. children in interventions targeted to parents, or individuals with chronic conditions in interventions targeted to caregivers) or focussed solely on the severity of a chronic physical condition (e.g. diabetes or cancer). Measurement tools were also excluded if they were bespoke

tools developed specifically for the study (e.g. not a pre-existing standardised measurement tool); were an existing tool that had been modified specifically for the study (e.g. changing questionnaire wording, question order, or overall length); were a subscale of an existing tool used in isolation of the full measure (e.g. use of only the anxiety subscale of a general symptom inventory). It has been suggested that these practices can reduce the transparency of measurement in a study and introduce researcher degrees of freedom.⁸

Systematic review and meta-analysis studies were also excluded, however, the reference lists of reviews identified by the search strategy and deemed relevant to mental health prevention were hand-searched for additional eligible records.

Record Screening

Initial title and abstract screening were conducted by a single researcher (VZ) supervised by JB. The full text of articles included from the initial screening were then screened according to the inclusion criteria by both VZ and a second independent reviewer (ES or EU). Conflicts were discussed within the study team, and if necessary, resolved by JB and EU.

Data Extraction and Synthesis

Data from included records were extracted by VZ and ES, with a subset of 10% of records reviewed for accuracy by a third reviewer (EU).

The following data were extracted from each included study: basic information about the study (author, year of publication, country, aim), information about the sample (age, gender, ethnicity), information about the intervention (intervention description, setting, primary and secondary outcomes) and information about measurement (measurement tools used for evaluation of outcomes). Each unique measurement tool identified during the initial extraction was listed and categorised by target construct (the outcome the tool is designed to measure). The target constructs were then coded by VZ into the following higher-order categories: symptoms, diagnosis, functioning, protective factors, risk factors, cognitive factors, quality of life, wellbeing, help-seeking, cost-effectiveness, and other. Given the focus of the study on the prevention of common mental health problems, measures categorised as measures

of mental health symptoms or diagnosis were selected for secondary extraction. Measures in the remaining categories were summarised descriptively.

Following identification of mental health measurement tools, secondary literature searches were conducted for papers providing further information about the characteristics (e.g. psychometrics, development, terms of use) of each tool. Extraction of measurement tool characteristics captured: basic information about the measure (developer, year, target population), psychometric properties (content, construct, and criterion validity, internal consistency, and test-retest reliability), practicability (number of items, costs, terms of use, administration method, availability of normative data), and sensitivity (cultural sensitivity). The checklist was developed by VZ and was pilot tested by ES and EU and revised following use on a sample of initial records.

Assessment of Quality

The criteria for evaluating the psychometric properties, and quality of the investigation of psychometric properties were drawn from the validity and reliability sections of the Medical Outcomes Trust Scientific Advisory Committee Guidance on Evaluating Quality of Life and Health Status Instruments.⁹ Evidence for cultural sensitivity was assessed using criteria similar to those used in a systematic review of subjective wellbeing measurement tools¹⁰ and was assessed as the availability of translations of the measure in languages other than English, or the validation or successful use of the measure with individuals from diverse cultures. Each characteristic was rated as “present”, “not present”, “mixed”, or “unknown”.

Systematic review of the literature regarding the properties of each individual measure was beyond the scope of the current study. Non-systematic literature searches were conducted for each tool. Where multiple research articles were available which assessed the psychometric properties of a measure, the recommendations outlined in Greenhalgh et al., (1998) were followed.¹¹ Preference was given to the record which described the original development of the tool, or the next closest record (in terms of year and author involvement) if the original record was inaccessible. Where a systematic review or meta-

analysis of psychometric properties was available, a critical reading of the review was used as evidence of psychometric properties.

Results

Identification of Prevention Evaluation Studies

Searches identified a total of 11,962 records, 1,911 duplicate records were excluded leaving 10,051 records eligible for initial screening. Of these, 9,222 records were found to be irrelevant at the title and abstract screening stage, leaving 829 full-text studies eligible for secondary screening.

Of these, a further 712 studies were deemed ineligible due to not meeting one or more of the inclusion criteria. Most studies were excluded due to prevention of common mental health problems not being the stated primary aim, or due to non-intervention study designs.

A total of 117 studies met inclusion criteria. Four of these records were published trial protocols for trials with results also included in full by the search criteria. As the trial protocol and the study findings will both contain information about the measurement tools used, the records for the published protocol and published finding were combined to avoid double-counting the use of measurement tools. This resulted in 113 unique studies. An additional 14 studies were identified by hand-searching reference lists from relevant systematic reviews and meta-analyses. Within these 127 studies (Figure 1), 65 unique mental health measurement tools (symptom inventories and diagnostic measures) were identified and 125 non-mental health specific measurement tools were identified (measures of functioning, protective and risk factors, cognitive factors, quality of life, wellbeing, help-seeking, cost-effectiveness and other).

Description of Included Studies

Characteristics of included studies are summarised in Table 1. Identified studies took place mainly in the USA or Canada (41%, n = 52), used selective approaches (targeted to groups or populations known to have higher risk) (54%, n = 69), and 25 % (n=32) were targeted toward pregnant women or new mothers.

Most approaches were based on cognitive-behavioural therapy (CBT) (21%, n = 27), followed by combined interventions comprising a mix of approaches (e.g. CBT, psychoeducation, peer support etc.)

(19%, n = 24), and exclusively psychoeducation (16%, n = 20). The remaining studies used other psychological approaches (e.g. problem-solving therapy) (n = 14), mindfulness (n = 9), physical exercise (n = 7), counselling (n = 6), skills-training (n = 5), creative approaches (e.g. singing or arts classes) (n = 4), and peer mentoring (n = 3). A total of eight studies (6%) used other approaches that included parenting interventions, web-based decision aids, and dietary changes.

Symptom change was used to evaluate the preventative intervention in 65% of studies (n = 83). Sixteen percent of studies (n = 20) included a type of diagnostic assessment (i.e. incidence, onset, or time-to-onset, or use of established 'cut-offs' or 'thresholds' on continuous symptom measures). There was overlap between these approaches and 19% (n = 24) measured both symptoms and diagnostic indicators as outcomes. In terms of the specific outcomes measured, depression (either symptoms or diagnosis) was an outcome in 75% of included studies (n = 95). This was followed by anxiety (39%, n = 49), overall mental health (i.e. psychological distress) (21%, n = 27), PTSD (9%, n = 12) and OCD (1%, n = 1). There was substantial overlap between these categories and many studies measured more than one outcome.

Description of Common Mental Health Measurement Scales

Included measurement tools were developed between 1953 and 2010 (median = 1994). Just over one third (34%, n = 22) were developed for use (or recommended by developers for use) in both clinical and non-clinical populations, 26% (n = 17) for the general population, 23% (n = 15) for clinical populations, and 14% (n = 9) for specific sub-groups such as new mothers, military members, trauma survivors, or individuals with specific chronic illnesses.

Sixty-five unique measurement tools which measured mental health symptoms or diagnosis were identified. Of these, 56 assessed symptoms and nine were diagnostic interviews. Of the symptom measurement tools, 18 focussed on overall mental health symptomatology (i.e. covered a range of symptoms from across categories of common mental health problems) and 20 focussed on depression. The remainder focussed on anxiety (n = 9), PTSD (n = 8) and OCD (n = 1).

The most commonly reported measure overall was the Centre for Epidemiologic Studies Depression Scale (CES-D)¹² which was used by 15.0% (n = 19) of studies. Most measurement tools identified (57%, n = 37) were only used by a single study. Use of measurement tools was more consistent when looking within specific mental health outcome categories, for example, within only those studies that included depression as an outcome, 20% (n = 19) used the CES-D. Outcome measurement was also more consistent when considering evaluation of the prevention of postnatal depression. The Edinburgh Postnatal Depression Scale (EPDS)¹³ is targeted toward pregnant women and new mothers and was one of the most commonly used tools both within this group (53.1%, n = 17) and across all included studies (13.4%, n = 17).

Twenty-three studies (18.1%) measured prevention by using a diagnostic interview, of which the Mini-International Neuropsychiatric Interview (MINI)¹⁴ (n = 9) and any version of the Structured Clinical Interview for DSM (SCID) (n = 9) were most commonly used.

The most commonly used tools to capture depression, anxiety, and general mental health are presented in Table 2. Their frequency of use is presented both across all included studies (n = 127) and within those studies that specified depression (n = 95), anxiety (n = 49), or general mental health (n = 27) as an outcome. As many studies measured more than one of these outcomes or utilised a general mental health measurement tool to capture outcomes of anxiety or depression, these categories are not mutually exclusive and percentages for use within categories may sum to more than 100%.

Psychometrics

Based on the literature reviewed to assess the psychometrics of the identified tools (original development papers or systematic reviews where available), included tools generally had adequate reliability and validity to assess mental health symptoms, and a majority (69.2%, n = 45) were validated and/or translated in multiple languages. The studies establishing psychometric properties of the tools were generally of mixed quality. Over half (58.5%) of identified tools (n = 38) met at least five of eight quality criteria, and 41.5% (n = 27) did not. This may reflect in part the pragmatic approach to reviewing the psychometric literature, in which original development papers, many of which are older publications

that may not reflect current standards in reporting, were used to assess psychometrics. The psychometric properties for the most frequently used tools are presented in Table 3.

For evaluation of depression prevention interventions, the CES-D and the Patient Health Questionnaire (PHQ-9)¹⁵ were the most widely used, freely available measurement tools. At 20 items, the CES-D is longer than the PHQ-9 (nine-items) and so may be more appropriate for research where participant burden (the demands placed on participant time as part of participation in a research study) is less of a concern. Both tools are available in a range of languages,^{16,17} are recommended by developers as appropriate for a general, non-clinical population, and have evidence to support their validity, internal consistency, and test-retest reliability.^{12,18,19} For anxiety, the seven-item Generalised Anxiety Disorder measure (GAD-7)²⁰ was the most widely used tool. The GAD-7 has been translated into multiple languages, is appropriate for use in a general, non-clinical population, and has evidence to support its validity, internal consistency, and test-retest reliability.^{16,20} For measurement of a range of mental health symptoms, the 21-item Depression, Anxiety, Stress Scale (DASS-21)²¹ was the most widely used, freely available, self-reported tool. It is appropriate for use with a general, non-clinical population and has evidence available to support its validity and internal consistency.^{22,23} It has been successfully used in a variety of cultures and countries and is available in multiple translated versions.²⁴

Practicability

The majority of tools are administered through self-report (73.8%, n = 48). We were unable to find clear details on terms of use for 29.2% (n = 19) of the scales. Of those with terms of use instructions available, half (50.0%, n = 23) were free-to-use, 45.7% (n = 21) had costs associated with use, and 4.3% (n = 2) varied between free use and paid use dependent on the type of work and funding source.

In addition to fees for use, 7.7% (n = 5) had restrictions on the type of work tools could be used for (e.g. not to be used for commercial gain), 7.7% (n = 5) had restrictions on the access and production of translated versions, 7.7% (n = 5) had restrictions on how the tool could be administered (e.g. paper copy

or online), and 3.1% (n = 2) required specific training prior to use. For 24.6% (n = 16) it was unclear if there were additional restrictions on use.

Description of Non-Mental Health Specific Measurement Tools

Across the 127 studies, a further 125 measurement tools were identified which measured outcomes other than symptoms or diagnoses of common mental health problems. Most outcome measurement tools classified as measuring “non-mental health specific” constructs, measured outcomes that are protective (n = 42) and risk factors (n = 28) presumed to contribute to increased or decreased risk of onset and/or increase of symptomatology.

Protective outcomes included measures of positive psychological constructs (e.g. compassion, hope, engagement, optimism, self-efficacy, and spirituality) as well as overall resilience, and behaviours or social and environmental factors which support good mental health (e.g. physical activity, job satisfaction, social support, coping-skills, and relationship quality).

Risk outcomes included overall measures of general stress and fatigue, as well as risk-behaviours (substance use) and environmental or social factors associated with increased risk of mental health problems (e.g. caregiving burden, parenting stress and practices, sleep quality, experience of trauma, exposure to domestic violence and abuse).

In addition to risk and protective outcomes, non-mental health specific outcome measures captured: cognitive factors (n = 18) (individual thought processes such as rumination, decision-making, problem-solving, or mindfulness), quality of life (n = 11), wellbeing (n = 8), general functioning (n = 8), help-seeking (n = 3), cost-effectiveness (n = 3) and other (mental health literacy, mental health stigma, and sexual functioning) (n = 4). It is beyond the scope of the current paper to further evaluate these measures; however, a separate piece of work is underway to capture these factors and assess their measurement properties.

Discussion

The present study provides an overview of the measures used to evaluate the effectiveness of interventions that prevent common mental health problems. It gives insight into the range of outcome measures used, and the ones most frequently used. This information is a first step in developing a potential future common outcomes framework for measuring the effect of preventative interventions in mental health.

Our findings suggest heterogeneity in measures used for evaluating the prevention of common mental health problems. Across 127 studies, 65 measurement tools were identified for measuring mental health problems and a further 125 were identified which measured non-mental health specific constructs. The most commonly reported tool, (CES-D), was only used in 19 out of 127 studies.

We identified several frequently used measures which were freely accessible with adequate psychometric properties. The most frequently used, freely accessible, tools included the CES-D and PHQ-9 for prevention of depression and the EPDS for prevention of depression in pregnant women or new mothers. The GAD-7 was the most frequently used freely available tool for assessing prevention of anxiety and the DASS-21 was the most frequently used, freely available, tool for assessing general mental health.

It should be noted that the PHQ-9 and GAD-7 are commonly used in clinical practice in England as part of the routine outcome monitoring component of the Improving Access to Psychological Therapies (IAPT) programme.²⁵ As such, use of the PHQ-9 and GAD-7 may have the additional benefit of providing results that are more directly comparable to clinical populations in England using routinely collected clinical data. It should be noted that the PHQ-9 includes an item regarding suicidal ideation, which can potentially raise safeguarding concerns beyond the scope of some community-based preventative mental health programmes to appropriately manage. In these instances, researchers may wish to consider the eight-item version, which was found to have broadly similar psychometric properties, but which excludes the item on suicidal ideation.¹⁸

The range of measurements chosen depend on the aim of the study, however, practical considerations also affect this decision. For instance, diagnostic assessments can help to accurately measure the impact

of a prevention programme on the incidence of common mental health problems like depression, especially when they are applied with a long follow-up. Only 23 studies (18.1%) measured prevention by using a diagnostic assessment tool in our review, of which, the Mini-International Neuropsychiatric Interview (MINI) and any version of the Structured Clinical Interview for DSM (SCID) were used most often. Part of the reason why these measures were not used more frequently may be because they commonly require training, expertise and costs for use. In these cases, a commonly used, free and validated self-reported tool may be more feasible, especially for voluntary and community sector organisations considering local evaluation of their work in prevention. Where budget is not constrained, diagnostic outcome assessments may be suitable, pending on the research question being asked.

To improve the evidence base for prevention, especially in applied settings, more consistency in measurement could benefit the field. Evidence suggests that, at a conceptual level, using measures with disparate specifications can create confusion when interpreting or comparing effectiveness across settings and population groups.²⁶ The use of adapted or bespoke measures may also further introduce researcher degrees of freedom to a study.⁸ Therefore, the evidence base may benefit from taking steps toward a common measurement framework that standardises and aligns evaluation of preventative interventions in the community. The present review represents a first step toward this goal. By providing an overview of the range of measurement tools currently used to evaluate mental health prevention interventions, we provide indication of the current degree of alignment and provide measurement tools that may be candidates for inclusion in future common measurement framework development.

Limitations

The study is limited by the lack of grey literature. Though searches were conducted in Open Grey, a database of grey literature, very few relevant results were identified. It is likely that there are a range of tools used to evaluate the impact of mental health prevention interventions outside of the academic literature which will have been missed from the present review. While many of these tools are likely to be bespoke to organisations or projects, it would be valuable to assess any existing standardised tools that are commonly used outside of the academic literature, particularly those used in voluntary and community sector settings. Furthermore, due to the number of tools included, it was not possible to

systematically review the evidence for the psychometric properties of each tool identified by this review. Instead, literature reviews were conducted which aimed to utilise critical readings of existing systematic reviews and meta analyses where available, and the original development paper, or next closest paper in terms of year and author involvement, as evidence of psychometric properties. This is a pragmatic approach which adheres to recommendations from Greenhalgh et al., (1998) for reviewing and selecting outcome measures for use in routine practice.¹¹ However, due to this, some relevant papers regarding the psychometrics for tools identified by this review may have been missed.

During the process of developing the search strategy and inclusion criteria for this review, it became clear that the boundaries between what is, and is not, a mental health prevention intervention can be unclear. This is something that has historically been a challenge in conceptualising prevention work in mental health.⁵ In order to apply consistent rules that achieved a manageable number of results we were only able to include studies with a stated primary aim of improving mental health in a subclinical population or preventing a common mental health problem. This aligns with the criteria suggested by Cowen (1980) for identifying and defining primary prevention work.²⁷ However, in doing so, it is possible that interventions which had broader and less clearly defined aims, but which were nonetheless relevant to the prevention of mental health problems, were excluded from review.

Implications for Future Research

While in-depth exploration of the secondary tools identified was beyond the scope of the current review, further exploration of the properties of the tools used to capture non-mental health specific outcomes would be beneficial for enhancing understanding of how current mental health prevention studies measure the secondary and intermediate outcomes of their interventions, and for developing recommendations for harmonisation of this aspect of measurement and evaluation.

In the absence of harmonisation of measurement of preventative interventions, it may be valuable to explore alternative strategies to allow movement between different mental health outcome measures, for example, the use of mapping approaches like those used to map the relationship between more generic quality of life measures and condition-specific measures.²⁸

Many of the commonly used tools identified were developed using items drawn from diagnostic criteria, or other validated questionnaires. To our knowledge, few of the most commonly used tools involved substantial co-production in the development of the items themselves, though several utilised consultation approaches (e.g. piloting and focus groups) to refine items. As such, it will be useful to assess the acceptability of tools identified by this review among those populations most likely to be participants in mental health prevention interventions. Understanding more about how participants view these scales, and their relevance to their own lives and circumstances, can provide support for tools being “participant-valued”.²⁹ Participant-valued tools are those that have been produced by researchers and clinicians using traditional methods but have been found to reflect the perspectives of participants through the use of qualitative research approaches.²⁹ The use of outcome measures that are patient-valued (and wherever possible, co-produced) is important to ensure that research accurately captures the aspects of mental health that are most important to the participants, and therefore has implications for both the meaningfulness, and the quality, of the data collected.²⁹ Plans for follow-up qualitative research of this kind are currently underway by the co-authors of the current study.

Funding

The present study was funded by the Mental Health Foundation.

Conflicts of Interest

No conflicts of interest to declare.

Acknowledgements

The authors would like to thank Isabella Goldie for offering her insight and expertise which assisted with developing the current research.

Key Points

- To build the evidence-base for mental health prevention, research should aim to improve standardisation of outcome measurement across studies

- Currently, there is little consistency across prevention studies in terms of the tools used to measure mental health prevention outcomes
- We propose that the most frequently used and freely available measurement tools identified by this review be considered when developing any future common measurement frameworks for the evaluation of mental health prevention approaches.

References

1. Friedrich MJ. Depression Is the Leading Cause of Disability Around the World. *JAMA*. 2017 Apr 18;317(15):1517.
2. Sansfeld S, Clark C, Bebbington P, King M, Jenkins R, Hinchcliffe S. Chapter 2: Common mental disorders. In: McManus S, Bebbington P, Jenkins R, Brugha T, editors. *Mental Health and wellbeing in England: Adult Psychiatric Morbidity Survey 2014*. Leeds: NHS Digital; 2016.

3. World Health Organisation. Depression and Other Common Mental Disorders Global Health Estimates. Geneva, Switzerland; 2017.
4. NICE. Common Mental Health Disorders in Primary Care Overview. 2016.
5. National Research Council (US) and Institute of Medicine (US) Committee on the Prevention of Mental Disorders and Substance Abuse Among Children and Young Adults: ; Y. Defining the Scope of Prevention. In: Research Advances and Promising Interventions. National Academies Press (US); 2009.
6. Goldie I, Elliot I, Regan M, Bernal L, Makurah L. Mental health and prevention: Taking local action for better mental health. 2016.
7. McDaid D, Park A-L, Knapp M, Ali N, Appleby L, Belsman L, et al. Commissioning Cost-Effective Services for Promotion of Mental Health and Wellbeing and Prevention of Mental Ill-Health. LSE Pers Soc Serv Res Unit. 2017;
8. Flake JK, Fried EI. Measurement schmeasurement: Questionable measurement practices and how to avoid them. Preprint. 2019;(January).
9. Lohr KN, Aaronson NK, Alonso J, Burnam MA, Patrick DL, Perrin EB, et al. Evaluating quality-of-life and health status instruments: development of scientific review criteria. *Clin Ther*. 1996;18(5):979–92.
10. Lindert J, Bain PA, Kubzansky LD, Stein C. Well-being measurement and the WHO health policy Health 2010: systematic review of measurement scales. *Eur J Public Health*. 2015 Aug 1;25(4):731–40.
11. Greenhalgh J, Long A, Brettle A, Grant M. Reviewing and selecting outcome measures for use in routine practice. *J Eval Clin Pract*. 1998 Nov 1;4(4):339–50.
12. Radloff LS. The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. *Appl Psychol Meas*. 1977 Jun 26;1(3):385–401.
13. Cox JL, Holden JM, Sagovsky R. Detection of Postnatal Depression: Development of the 10-item Edinburgh Postnatal Depression scale. *Br J Psychiatry*. 1987 Jun 2;150(JUNE):782–6.
14. Sheehan D V, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry*. 1998;59 Suppl 2:22-33;quiz 34-57.
15. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. 2001 Sep;16(9):606–13.
16. PHQ and GAD-7 Instructions INSTRUCTION MANUAL Instructions for Patient Health Questionnaire (PHQ) and GAD-7 Measures.
17. Roberts RE, Vernon SW, Rhoades HM. Effects of language and ethnic status on reliability and validity of the Center for Epidemiologic Studies-Depression Scale with psychiatric patients. *J Nerv Ment Dis*. 1989 Oct;177(10):581–92.
18. Kroenke K, Spitzer RL, Williams JBW, Löwe B. The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: A systematic review. *Gen Hosp Psychiatry*. 2010 Jul;32(4):345–59.

19. Shafer AB. Meta-analysis of the factor structures of four depression questionnaires: Beck, CES-D, Hamilton, and Zung. *J Clin Psychol*. 2006 Jan;62(1):123–46.
20. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: The GAD-7. *Arch Intern Med*. 2006 May 22;166(10):1092–7.
21. Lovibond S. Manual for the depression anxiety stress scales. 2nd ed. Sydney N.S.W.: Psychology Foundation of Australia; 1995.
22. Sinclair SJ, Siefert CJ, Slavin-Mulford JM, Stein MB, Renna M, Blais MA. Psychometric Evaluation and Normative Data for the Depression, Anxiety, and Stress Scales-21 (DASS-21) in a Nonclinical Sample of U.S. Adults. *Eval Heal Prof*. 2012 Sep 18;35(3):259–79.
23. Antony MM, Cox BJ, Enns MW, Bieling PJ, Swinson RP. Psychometric properties of the 42-item and 21-item versions of the Depression Anxiety Stress Scales in clinical groups and a community sample. *Psychol Assess*. 1998;10(2):176–81.
24. Norton PJ. Depression Anxiety and Stress Scales (DASS-21): Psychometric analysis across four racial groups. *Anxiety, Stress Coping*. 2007 Sep;20(3):253–65.
25. National Collaborating Centre for Mental Health. *The Improving Access to Psychological Therapies Manual*. 2018;1–15.
26. National Quality Forum. *Guidance for Measure Harmonization: A Consensus Report*. Washington, DC; 2010.
27. Cowen EL. The wooing of primary prevention. *Am J Community Psychol*. 1980 Jun;8(3):258–84.
28. Brazier J, Connell J, Papaioannou D, Mukuria C, Mulhern B, Peasgood T, et al. A systematic review, psychometric analysis and qualitative assessment of generic preference-based measures of health in mental health populations and the estimation of mapping functions from widely used specific measures. *Health Technol Assess (Rockv)*. 2014 May;18(34):1–188.
29. Trujols J, Portella MJ, Iraurgi I, Campins MJ, Siñol N, Cobos JP de L. Patient-reported outcome measures: Are they patient-generated, patient-centred or patient-valued? Vol. 22, *Journal of Mental Health*. 2013. p. 555–62.

Figure 1: Study flow diagram for systematic review process

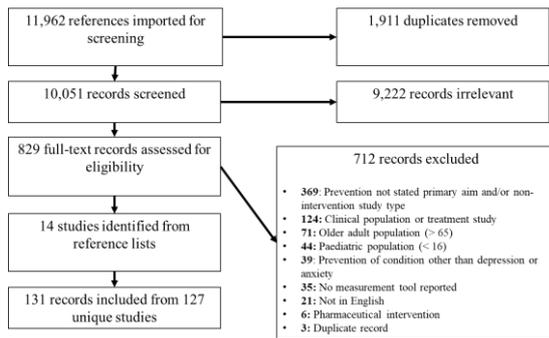


Table 1: Characteristics of included studies

	N	%
<hr/>		
Country		
UK	17	13.4
Other European	24	18.9
USA & Canada	52	40.9
Australia & New Zealand	21	16.5
Asia	9	7.1
Other	4	3.1
Preventative Approach		
Universal	27	21.3
Selective	69	54.3
Indicated	31	24.4
Population Group		
General Population	29	22.8
Pregnant Women/New Mothers	32	25.2
Students and Young Adults	23	18.1
Chronic Physical Illness/Disability	13	10.2
Families and Caregivers	14	11.0
Trauma Survivors	5	3.9
Other	11	8.7

Table 2: Most commonly used measurement tools by outcome

	Number of Uses	% of all studies	% of studies with specified outcome
Depression		(n = 127)	(n = 95)
*CES-D	19	15.0	20.0
*EPDS	17	13.4	17.9
BDI-II	16	12.6	16.8
*PHQ-9	12	9.4	12.6
Anxiety		(n = 127)	(n = 49)
*GAD-7	7	5.5	14.3
BAI	4	3.1	8.2
*PSWQ	4	3.1	8.2
STAI	4	3.1	8.2
General Mental Health		(n = 127)	(n = 27)
*DASS-21	12	9.4	44.4
HADS	9	7.1	33.3
*PANAS	5	3.9	18.5
GHQ-12	4	3.1	14.8

*tool is in public domain, or free to use for non-profit research/clinical purposes

Table 3: Psychometric properties of most commonly used measurement tools

	Construct	Criterion	Internal	Test-Retest
	Validity	Validity	Consistency	Reliability
Depression				
*CES-D	+	+/-	+	+/-
BDI-II	+	+/-	+	+
*EPDS	+	+	\	\
*PHQ-9	+	+	+	+
Anxiety				
*GAD-7	+	+	+	+
BAI	+	+	+	+
*PSWQ	+	+	+	+
STAI	+/-	\	+	+/-
General Mental Health				
*DASS-21	+	\	+	\
HADS	+	+	+	\
*PANAS	+	+	+	+/-
GHQ-12	+	+	+	+

*tool is in public domain, or free to use for non-profit research/clinical purposes;

+, good psychometric values; -, less than optimal psychometric values; ±, not consistent value; \ unknown or unclear