International consensus on a standard set of outcome measures for child and youth anxiety, depression, obsessive-compulsive disorder, and post-traumatic stress disorder

Karloin R Krause, Sophie Chung, Abiadun O Adewuya, Anne Marie Albano, Rochelle Babins-Wagner, Laura Birkinhshaw, Peter Braron, Cathy Creswell, Kathleen Delaney, Bruno Falissard, Christopher B Forrest, Jennifer L Hudson, Shin-ichi Ishikawa, Meghna Khatwani, Christian Kording, Judi Krause, Kanika Mallik, Vania Martinez, Fanzo Moghal, Thomas H Ollendick, Say How Ong, George C Patton, Ulrike Rawnsley-Sieherer, Peter Sazitmari, Evie Thomas, Lucas Walters, Bryan Young, Yue Zhao, Miranda Wolpert

A major barrier to improving care effectiveness for mental health is a lack of consensus on outcomes measurement. The International Consortium for Health Outcomes Measurement (ICROM) has already developed a consensus-based standard set of outcomes for anxiety and depression in adults (including the Patient Health Questionnaire-9, the Generalised Anxiety Disorder 7-item Scale, and the WHO Disability Schedule). This Position Paper reports on recommendations specifically for anxiety, depression, obsessive-compulsive disorder, and post-traumatic stress disorder in children and young people aged between 6 and 24 years. An international ICHOM working group of 27 clinical, research, and lived experience experts formed a consensus through teleconferences, an exercise using an adapted Delphi technique (a method for reaching group consensus), and iterative anonymous voting, supported by sequential research inputs. A systematic scoping review identified 70 possible outcomes and 107 relevant measurement instruments. Measures were appraised for their feasibility in routine practice (ie, brevity, free availability, validation in children and young people, and language translation) and psychometric performance (ie, validity, reliability, and sensitivity to change). The final standard set recommends tracking symptoms, suicidal thoughts and behaviour, and functioning as a minimum through seven primarily patient-reported outcome measures: the Revised Children’s Anxiety and Depression Scale, the Obsessive Compulsive Inventory for Children, the Children’s Revised Impact of Events Scale, the Columbia Suicide Severity Rating Scale, the KIDSCREEN-10, the Children’s Global Assessment Scale, and the Child Anxiety Life Interference Scale. The set’s recommendations were validated through a feedback survey involving 487 participants across 45 countries. The set should be used alongside the anxiety and depression standard set for adults with clinicians selecting age-appropriate measures.

Introduction

Depression and anxiety affect an estimated 4-4% and 3-6% of the world’s population, and rank as the first and sixth largest contributor to health-related disability, respectively. These disorders frequently emerge in childhood and adolescence and, unless treated early and effectively, commonly adversely affect mental health and psychosocial outcomes across the life course. Despite an increase in mental health-care provision, service systems have failed to reduce the prevalence of these disorders in children and young people. The global response to this problem requires holistic strengthening, not only in specialist mental health services, but also in primary care, community health, child health, and school settings.

In addition to resourcing and training in evidence-based care, one essential element of service strengthening is the systematic monitoring of patient progress. Valid data for treatment outcomes are an essential facet for evaluating the effectiveness of care and can inform the setting of strategic targets for health systems, comparisons between different systems or services, and clinical decision making on a case-by-case basis. Currently, there is neither agreement nor global guidance on how best to track the response to clinical care for anxiety and depression in children and young people. Uptake of routine outcome measurement is low and, when done, there is considerable variation in outcomes, instruments, and assessment timepoints; a review published in 2019 recorded 15 different measures used to assess primary outcomes across 19 studies of routine treatment for anxiety and depression in young people. Resulting data gaps and inconsistencies severely limit the potential for comparing different models of clinical care, identifying good practice, and informing quality improvement efforts.

This initiative aimed to address this challenge by devising a standard set, that is, a consensus-based standardised collection of treatment outcomes to be measured and reported as a minimum by all those providing relevant care. To ensure that this standard is meaningful and acceptable to its intended users, the International Consortium for Health Outcomes Measurement (ICHOM) convened an international working group of lived experience experts, practitioners, and researchers to agree on a set of outcome domains, measurement instruments, case-mix factors (ie, case characteristics that should be considered when adjusting for differences in the composition of service user populations, or care provision across settings), and measurement timepoints (panel 1).

An existing ICHOM standard set for adult anxiety and depression covers young people from the age of 14 years (appendix p 3). It includes a recommendation for tracking symptom change via the Generalised Anxiety
Disorder 7-item Scale (GAD-7)\(^1\) and the depression subscale of the Patient Health Questionnaire (PHQ-9),\(^2\) as well as functioning through the WHO Disability Assessment Schedule (WHODAS) 2.0 12-item short form.\(^3\) Our standard set aims to complement this effort by providing recommendations specifically tailored for use with children and young people. The combination of the two sets will provide for the transition from youth into adult services at any point between the ages of 14 and 24 years, allowing for local variation in transition and judgments about which set is most suitable for different ages.

This standard set is designed for children and young people aged between 6 and 24 years who access care for anxiety, depression, obsessive-compulsive disorder (OCD), or post-traumatic stress disorder (PTSD), as defined by standard diagnostic criteria. All are internalising disorders, typically characterised by high negative affect, with OCD and PTSD long classified as anxiety disorders.\(^4\) This standard set is recommended for use by all those providing care to the population in scope worldwide, regardless of intervention setting or approach. The working group sought to combine self-reported, parent-reported, and clinician-reported outcome measures to account for the different perspectives these reporters can provide.\(^5\)–\(^8\) Parent report measures also serve as a proxy when children and young people are unable to complete measures because of young age or developmental constraints (appendix p 3).\(^9\)

The process and methods used to develop this standard set are summarised in panel 2 and shown in figure 1, with additional detail provided in the appendix (pp 3–18).

**Recommended outcomes and measures**

As per the working group consensus, this standard set recommends tracking response to treatment across the three outcome domains of symptoms, suicidal thoughts and behaviour, and functioning, as a minimum, using seven instruments that are primarily self-reported. These seven measurement instruments were selected because they fulfilled most or all of the working group’s appraisal criteria (table 1, panel 2). A detailed discussion of instrument properties, performance against appraisal criteria, and accessibility is provided in the appendix (pp 19–22). 463 practitioners and researchers, and 24 lived-experience experts provided feedback on the draft working group recommendations through an open review process (panel 2, appendix p 18). Overall, 75% of practitioners and researchers stated that they had confidence in the recommended outcomes and instruments. In addition, 75–100% of participants with lived experience rated each included outcome domain as important, and 100% of lived experience participants confirmed the acceptability of the recommended measures for use in clinical practice. No additional outcome domains were consistently highlighted as missing from the recommended set during the open review process.

### Symptoms

The set recommends measuring anxiety and depression symptoms through youth and parent report for all children and young people within the prespecified scope (ie, those aged between 6 and 24 years presenting with anxiety, depression, OCD, or PTSD). To minimise the length and complexity of the standard set, the working group chose to recommend a joint measure of anxiety and depression symptoms—the 25-item short form of the Revised Children’s Anxiety and Depression Scale (RCADS-25).\(^10\) The group initially selected the scale’s 47-item version on the basis of the appraisal criteria,\(^11\) but decided to recommend the short form to reduce respondent burden. The RCADS-25 consists of 15 items that track anxiety symptoms and ten items that track symptoms of major depression, which can be summed to compute aggregate anxiety, depression, and total internalising symptom scores. Although less widely validated than the RCADS-47, the RCADS-25 met most inclusion criteria (table 1), although evidence of its sensitivity to change was not available at the time of the appraisal. Long or short versions of the RCADS have been applied in Africa, Europe, the Americas, and Asia.\(^12\)

Symptoms of OCD and PTSD should be tracked separately for children and young people presenting with a diagnosis of OCD or PTSD, or with sub-threshold symptoms as appropriate. Symptoms of OCD and PTSD should be tracked via the self-reported 21-item Obsessive

---

**Panel 1: The International Consortium for Health Outcomes Measurement (ICOM) approach to standard set development**

ICOM is a non-profit organisation that specialises in the development of condition-specific standard sets for clinical practice. ICOM has supported the development of more than 28 existing standard sets, including one for adult depression and anxiety.\(^13\) Outcome measurement is approached within a framework of person-centred and value-based health care, in which value is defined as the health outcomes achieved, relative to the resources invested, rather than the volume of services delivered.\(^14\) Within a person-centred framework, value should be defined around outcomes that matter to service users. All ICHOM standard sets are condition specific, based on the understanding that the needs of service users and the treatment options available to them are at least partly shaped by the principal presenting problems.\(^15\) Service users are directly involved in defining the standard set, which must include patient-reported outcomes. The final set of outcomes should represent the end result rather than the process of care, should balance a comprehensive approach to tracking outcome with a feasible recommendation that services can reliably implement, and should be responsive to quality improvement efforts.

---

**ICOM**

The International Consortium for Health Outcomes Measurement (ICHOM) approach to standard set development:

The International Consortium for Health Outcomes Measurement (ICHOM) approach to standard set development:

The International Consortium for Health Outcomes Measurement (ICHOM) approach to standard set development:

The International Consortium for Health Outcomes Measurement (ICHOM) approach to standard set development:

The International Consortium for Health Outcomes Measurement (ICHOM) approach to standard set development:

The International Consortium for Health Outcomes Measurement (ICHOM) approach to standard set development:

The International Consortium for Health Outcomes Measurement (ICHOM) approach to standard set development:

The International Consortium for Health Outcomes Measurement (ICHOM) approach to standard set development:

The International Consortium for Health Outcomes Measurement (ICHOM) approach to standard set development:

The International Consortium for Health Outcomes Measurement (ICHOM) approach to standard set development:

The International Consortium for Health Outcomes Measurement (ICHOM) approach to standard set development:

The International Consortium for Health Outcomes Measurement (ICHOM) approach to standard set development:

The International Consortium for Health Outcomes Measurement (ICHOM) approach to standard set development:

The International Consortium for Health Outcomes Measurement (ICHOM) approach to standard set development:

The International Consortium for Health Outcomes Measurement (ICHOM) approach to standard set development:

The International Consortium for Health Outcomes Measurement (ICHOM) approach to standard set development:

The International Consortium for Health Outcomes Measurement (ICHOM) approach to standard set development:

The International Consortium for Health Outcomes Measurement (ICHOM) approach to standard set development:

The International Consortium for Health Outcomes Measurement (ICHOM) approach to standard set development:

The International Consortium for Health Outcomes Measurement (ICHOM) approach to standard set development:

The International Consortium for Health Outcomes Measurement (ICHOM) approach to standard set development:

The International Consortium for Health Outcomes Measurement (ICHOM) approach to standard set development:

The International Consortium for Health Outcomes Measurement (ICHOM) approach to standard set development:

The International Consortium for Health Outcomes Measurement (ICHOM) approach to standard set development:

The International Consortium for Health Outcomes Measurement (ICHOM) approach to standard set development:

The International Consortium for Health Outcomes Measurement (ICHOM) approach to standard set development:

The International Consortium for Health Outcomes Measurement (ICHOM) approach to standard set development:

The International Consortium for Health Outcomes Measurement (ICHOM) approach to standard set development:

The International Consortium for Health Outcome...
Panel 2: Methods

The working group comprised 27 experts from 13 countries, including youth with lived experience, parents or carers (hereafter referred to as parents), mental health practitioners, and researchers working within relevant disciplines (eg, psychiatry, epidemiology, psychometrics; appendix pp 3–4). The central project team (KRK, SC, MW) coordinated and facilitated the consensus-building process and completed supporting research tasks but did not vote on the consensus recommendations. A detailed description of the process and methods is provided in the appendix (pp 3–18).

General process

From October, 2018, until December, 2019, the working group completed a 14-month structured and evidence-informed consensus-building process (figure 1). The group convened for eight teleconferences, completed a three-round adapted Delphi exercise (a method for reaching group consensus) to select outcome domains,10,11 and participated in iterative rounds of anonymous voting to arrive at recommendations for outcomes, measurement instruments, case-mix factors, and timepoints (appendix pp 4–5). The process was informed by sequential research inputs done by the central project team (appendix pp 5–18).

Selection of outcome domains

In line with methodological recommendations by the Core Outcome Measures in Effectiveness Trials (COMET) initiative,14 70 possible outcomes and 507 measurement instruments were identified through multiple avenues, including a systematic scoping review of 257 treatment outcome studies, and supplemental sources (eg, cohort studies, qualitative outcome research, and instrument banks), and break-out groups with user representatives (appendix pp 5–8).

Selection of measurement instruments

Once the group reached a consensus on the outcomes to include, 107 thematically relevant measures were systematically appraised to identify those most suitable for tracking the selected outcomes over time. Appraisal criteria included relevance (ie, comprehensive coverage of the selected outcome domain), feasibility and acceptability (ie, completion within less than 20 min, free availability for use in clinical settings, including in paper-and-pencil format, previous validation in children and young people, and translation into at least a second language), and psychometric performance (ie, inter-rater or retest reliability above 0·70, internal consistency above 0·70, and evidence of sensitivity to change), in line with International Society for Quality of Life Research recommendations (appendix pp 8–16).15 The instruments judged to best satisfy these criteria were shortlisted, and their strengths and shortcomings discussed. At this stage, other aspects, such as content and construct validity, were also considered. The final set of measures was then selected via consensus through iterative rounds of anonymous voting. The working group aimed for a standard set that would be simple to use, impose a low burden on its intended users, and be applicable across different contexts.

Selection of case-mix factors and measurement timepoints

Based on the systematic scoping review and existing ICHOM standard sets, the central project team compiled a list of possible case-mix factors, and did a rapid review of reviews examining predictors, mediators, and moderators of treatment response (appendix pp 17–18). The working group discussed this information and formed consensus on the case-mix factors to include. Similarly, the central project team drew on existing ICHOM standard sets to present an initial proposal for measurement timepoints, which were discussed, refined, and voted on by the working group.

Open review of draft recommendations

To establish the generalisability and acceptability of the working group’s recommendations, an open review web survey gathered external feedback from 463 practitioners, researchers, and policy makers from 45 countries, as well as 24 young people and parents from Denmark, the UK, and the USA (appendix p 18).
Functioning

Functioning describes a child’s ability to engage in typical activities and meet role demands in line with age-specific sociocultural norms. None of the identified functioning measures fully satisfied the working group’s requirements and a consensus was reached on mitigating this issue by tracking a broad concept of global functioning or health-related quality of life (HRQoL), as well as condition-specific functional impairment through short dedicated measures of each concept. Generic measures allow for comparisons across conditions, although condition-specific measures might be more sensitive to change. As per group consensus, measures had to cover psychosocial functioning, peer relationships, and sleep functioning at least at an item level, although sleep was eventually covered through the RCADS-25 (ie, item 8 (“I have trouble sleeping”) and item 9 (“I feel scared if I have to sleep on my own”)).

The KIDSCREEN-10 (completed by children or young people, and parents) was selected as a generic measure of global functioning. This unidimensional ten-item index of HRQoL tracks functioning in relation to physical health and energy levels, leisure activities, social and family relationships, and cognition. The more comprehensive KIDSCREEN-52 was originally developed through a process of cross-cultural harmonisation involving 13 European countries. A simultaneous approach was used, taking into account different cultural perspectives, to ensure international comparability and cross-cultural applicability. Its ten-item short form has been applied in Asia, eastern Africa, Europe, North America, and Latin America. Although originally designed for epidemiological studies, the KIDSCREEN-10 has been shown to discriminate well between children and young people with high and low levels of functioning, with few so-called ceiling or floor effects. In addition to the KIDSCREEN-10, the Children’s Global Assessment Scale (CGAS) was selected as a brief clinician-rated measure of global functioning.

On this widely used measure, clinicians perform a single rating by locating children and young people on a scale from 1 to 100, placing them into one of ten categories, rating by locating children and young people on a scale from 1 to 10 (indicating high impairment) to 91 to 100 (indicating superior functioning). The Children’s Anxiety Life Interference Scale (CALIS) was selected to measure condition-specific impairment via nine or ten items that describe instances of anxiety that a group did not identify any eligible measure that tracked depression-specific impairment or that captured impairment from both anxiety and depression. However, as of March, 2020, the CALIS author team were revising the measure to cover impairment from both anxiety and depression, and validation studies involving children, young people, and parents in community, school, and clinical settings were ongoing. As soon as validation results and the revised measure become available the working group will consider replacing the original CALIS with the revised Children’s Anxiety and Depression Life Interference Scale (CADLIS) in the standard set.

Figure 1: Working group process

<table>
<thead>
<tr>
<th>Break-out groups</th>
<th>Working group process</th>
<th>Research and external input</th>
</tr>
</thead>
<tbody>
<tr>
<td>Four thematic break-out groups to review shortlisted measures by outcome domain</td>
<td>October, 2018</td>
<td>Systematic scoping review of outcomes, measures, and case-mix factors</td>
</tr>
<tr>
<td></td>
<td>January, 2019</td>
<td>562 articles screened</td>
</tr>
<tr>
<td></td>
<td>March, 2019</td>
<td>257 articles included†</td>
</tr>
<tr>
<td></td>
<td>May, 2019</td>
<td>Outcomes/70 domains identified</td>
</tr>
<tr>
<td></td>
<td>June, 2019</td>
<td>Measures/50 unique measures identified</td>
</tr>
<tr>
<td></td>
<td>October, 2019</td>
<td>107 relevant measures appraised for feasibility</td>
</tr>
<tr>
<td></td>
<td>November, 2019</td>
<td>31 measures retained for psychometric assessment using ISOQOL criteria†</td>
</tr>
<tr>
<td></td>
<td>December, 2019</td>
<td>Open review process/46 professionals</td>
</tr>
<tr>
<td></td>
<td></td>
<td>24 lived-experience experts</td>
</tr>
</tbody>
</table>

Consolidation of discussion points in feedback survey and online voting on key decisions about scope, measures, case-mix, and timetables

Consolidation of discussion points and Delphi exercise to select outcome domains for inclusion

Position Paper
### Overview of recommended outcome measures and their evaluation criteria, as of March, 2020

<table>
<thead>
<tr>
<th>Measure</th>
<th>Relevance (domain coverage [ie, satisfactory domain coverage at subscale or item level])</th>
<th>Feasibility and acceptability</th>
<th>Psychometric performance*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(Short ie, &lt;60 items, &gt;20 min) Free ie, currently no cost for non-commercial use</td>
<td>Validated [ie, at least one validation in a CYP sample] Translated [ie, &gt;2 language version available]</td>
<td>Reliability [ie, TRT or IRR ≥0.70] Validity [ie, internal consistency [eg, Cronbach’s α&gt;0.70]] Sensitivity to change [ie, any evidence of sensitivity to change]</td>
</tr>
<tr>
<td><strong>Symptoms</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anxiety and depression</td>
<td>RCADS-25 GAD, MDD, OCD, PD, SAD, SP</td>
<td>25 items (&gt;10 min) Yes</td>
<td>Ages 6–18 years (clinical and non-clinical) TRT ≥ Yes*** No evidence</td>
</tr>
<tr>
<td>OCD</td>
<td>OCI-CV Doubting or checking, obsessing, hoarding, washing, ordering, neutralising</td>
<td>21 items (&gt;10 min) Yes</td>
<td>Ages 6–18 years (clinical and non-clinical) TRT ≥ Yes*** Some evidence†</td>
</tr>
<tr>
<td>PTSD</td>
<td>CRIES-8 (for children and youth) and CRIES-13 (for parent report) Intrusion, avoidance (hyperarousal in CRIES-13)</td>
<td>8 items (CRIES-8) and 13 items (CRIES-13); (&lt;5 min) Yes</td>
<td>Ages 7–18 years (clinical and non-clinical) TRT ≥ Yes*** Some evidence††</td>
</tr>
<tr>
<td><strong>Suicidal thoughts and behaviour</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Measure</td>
<td>C-SSRS</td>
<td>Severity of ideation, behaviour (ie, attempts) 3 or 6 items (&lt;5 min) Yes</td>
<td>Ages 12–18 years (clinical and non-clinical) &gt;100 IRR Yes†† Some evidence*</td>
</tr>
<tr>
<td><strong>Functioning</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Measure</td>
<td>KIDSCREEN-10 Physical activity and energy, emotions, leisure-time and participation, relationships with parents and peers, cognition, school performance</td>
<td>10 items (&lt;5 min) Yes</td>
<td>Ages 8–18 years (non-clinical) TRT ≥ Yes*** No evidence</td>
</tr>
<tr>
<td>Measure</td>
<td>CGAS</td>
<td>Global functioning</td>
<td>1 item (&lt;5 min) Yes</td>
</tr>
<tr>
<td>Measure</td>
<td>CALIS Enjoyable activities, relationships with siblings, parents, friends, peers, sports, schoolwork, distress</td>
<td>9 or 10 items (&lt;5 min) Yes</td>
<td>Ages 6–17 years (clinical and non-clinical) 7 TRT Yes** No evidence</td>
</tr>
</tbody>
</table>

CALIS=Child Anxiety Life Interference Scale. CGAS=Children’s Global Assessment Scale. CRIES-8=Children’s Revised Impact of Events Scale for youth. CRIES-13=Children’s Revised Impact of Events Scale for clinical and non-clinical. C-SSRS=Columbia Suicide Severity Rating Scale. CYP=children and young people. GAD=generalised anxiety disorder. MDD=major depressive disorder. NA=not applicable. OCD=obsessive compulsive disorder. OCI-CV=Obsessive Compulsive Inventory for Children. PD=panic disorder. RCADS-25=Revised Children’s Anxiety and Depression Scale. SAD=separation anxiety disorder. SP=social phobia. TRT=test-retest reliability. ǂThese thresholds are based on the International Society for Quality of Life Research minimum standards for patient-reported outcome measures.§Two initial questions serve as screeners, with the three remaining severity items schedule were considered instead.

### Recommended case-mix factors

The standard set aims to facilitate comparisons and benchmarking of outcomes, which requires the collection of additional data to adjust for variation in populations and intervention settings. As per the working group consensus, services should record demographic, clinical, complexity, and intervention factors through a mix of self-report, parent report, and clinician-report (table 2). Beyond the working group, the case-mix factors that were suggested were endorsed by more than 80% of practitioners and researchers who provided feedback during the open review process. These factors represent a minimum that should be assessed by all those providing relevant care, and services might wish to add other indicators to meet local information needs.

### Demographic factors

The set recommends recording age, sex assigned at birth, gender, ethnic minority status, socioeconomic status, and the child’s living situation. Socioeconomic status should be measured by recording the highest level of education completed by any of the parents of children and young people as a widely accepted proxy that can be mapped onto the International Standard Classification of Education for international comparisons. The set includes one question about children and young people’s living situation and two questions that capture ethnic
minority and marginalised group status via self-report (table 2).

**Clinical factors**

Several studies show that symptom burden, symptom duration, and the presence of comorbidities affect treatment response in children and young people with depression or anxiety (appendix pp 17–18). The group recommends recording principal and comorbid presenting problems by administering the 30 problem descriptions of the current view tool. Although not equivalent with formal diagnoses, these problem descriptions broadly align with the diagnostic categories of the ICD-11 for paediatric populations. ICHOM standards are available for recording symptom duration and previous service use (table 2).

**Complexity factors**

Research suggests that parental mental health influences treatment response in children and young people with anxiety and depression (appendix pp 17–18). The set includes two questions about experiences with or diagnoses of mental health problems in the immediate family, and previous use of mental health services by the reporting parent. Although evidence is scarce and conflicting on the influence of adverse experiences (eg, childhood maltreatment) on treatment response, the set recommends recording trauma history via the current view tool. Here, clinicians can indicate a range of adverse experiences based on available information. Problems at school or work should be tracked as additional complexity factors, using the current view tool.

**Intervention factors**

Services should collect information about the intervention approach, including the intervention focus (ie, who is actively involved), treatment modality and prescribed medication (ie, as per lists of options compiled by the working group and through open review feedback), and intervention setting (ie, whether or not treatment was delivered via a digital platform or inpatient care, as opposed to other settings). Services might wish to record additional detail on intervention characteristics to meet local, regional, or national information needs.

**Recommended measurement timepoints**

Timelines for measurement are highly practical considerations and are likely to vary substantially across services. The standard set makes a minimum recommendation (figure 2) for measuring outcomes over the full cycle of care but encourages services to do so as often as is clinically helpful to inform decision making, or to align with the collection of local or national data. The suggested timepoints were widely endorsed by over 80% of practitioners and researchers participating in the open review process.

---

**Table 2: Case-mix factors in the standard set for child and youth anxiety, depression, OCD, and PTSD**

<table>
<thead>
<tr>
<th>Demographic factors</th>
<th>Description</th>
<th>Reporter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age*</td>
<td>Year of birth</td>
<td>CYP or parent</td>
</tr>
<tr>
<td>Sex*</td>
<td>Sex assigned at birth</td>
<td>CYP or parent</td>
</tr>
<tr>
<td>Gender identity</td>
<td>“Do you think of yourself as...?”</td>
<td>CYP</td>
</tr>
<tr>
<td>Parent or carer education*</td>
<td>Highest level of education completed by any of the parents or carers of a child or young person (using ISCED standards)</td>
<td>CYP or parent</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>“Do you consider yourself to be in an ethnic minority where you live?”</td>
<td>CYP or parent</td>
</tr>
<tr>
<td>Marginalised group status</td>
<td>“Do you consider yourself to be a member of a marginalised group where you live?”</td>
<td>CYP or parent</td>
</tr>
<tr>
<td>Living situation</td>
<td>“Which of the following people live with you (your child) at your [their] home?”</td>
<td>CYP or parent</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Clinical factors</th>
<th>Description</th>
<th>Reporter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnoses and comorbidities</td>
<td>Measured via the provisional problems list of the current view tool</td>
<td>Clinician</td>
</tr>
<tr>
<td>Duration of symptoms*</td>
<td>“For how many months have you (your child) been experiencing [specific condition] symptoms?”</td>
<td>CYP or parent</td>
</tr>
<tr>
<td>Previous service use</td>
<td>“During the last year, did you (your child) receive any of the following treatments for [specific condition]? Complete separately for (a) medication, (b) psychotherapy, and (c) other.”</td>
<td>CYP or parent</td>
</tr>
</tbody>
</table>

**Complexity factors**

<table>
<thead>
<tr>
<th>Description</th>
<th>Reporter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trauma history</td>
<td>Measured via the selected complexity factors of the current view tool</td>
</tr>
<tr>
<td>Parental mental health</td>
<td>“It is useful to know whether there is a family history of mental health problems. Have you, or anyone else in the immediate family, ever experienced, or been diagnosed with, any of the following conditions: anxiety, depression, a substance use disorder (eg, alcohol or drugs), schizophrenic disorder, personality disorder, somatoform disorder (ie, unexplained physical symptoms), or other?”</td>
</tr>
<tr>
<td>Parental service use</td>
<td>“Have you ever sought help for your mental health?”</td>
</tr>
<tr>
<td>Education or work difficulties</td>
<td>Measured via the contextual problems section of the current view tool</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Intervention variables</th>
<th>Description</th>
<th>Reporter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervention focus (1)</td>
<td>“Who is actively involved in the intervention? (select all that apply)”</td>
<td>Clinician</td>
</tr>
<tr>
<td>Intervention focus (2)</td>
<td>“Is the intervention delivered to an individual child or family, or to a group of children or families?”</td>
<td>Clinician</td>
</tr>
<tr>
<td>Intervention approach</td>
<td>“What is the treatment approach? (select all that apply)”</td>
<td>Clinician</td>
</tr>
<tr>
<td>Prescribed medication</td>
<td>“What type of medication is prescribed? (select all that apply)”</td>
<td>Clinician</td>
</tr>
<tr>
<td>Intervention setting (1)</td>
<td>“Does this intervention involve an overnight stay at an institution providing mental health support?”</td>
<td>Clinician</td>
</tr>
<tr>
<td>Intervention setting (2)</td>
<td>“Does this intervention involve the use of a digital platform?”</td>
<td>Clinician</td>
</tr>
</tbody>
</table>

CYP=children and young people. ICHOM=International Consortium for Health Outcomes Measurement. ISCED=International Standard Classification of Education. OCD=obsessive-compulsive disorder. PTSD=post-traumatic stress disorder. *Variables defined and operationalised as per an existing ICHOM standard. Response categories for each question are provided in the full standard set reference guide (panel 3).

---

As per working group consensus, all case-mix factors and outcomes should be measured at the point of assessment or intake (ie, baseline), or as near to these timepoints as possible if services wish to collect data at second contact (figure 2). As a guideline, all outcome measures should be administered at least every 3 months after baseline measurements were taken. Services are encouraged to consider more frequent intervals, including session by session measurement, to help
Outcomes only
- Outcomes and case-mix factors

Active intervention(s)
- Collection at least every 3 months during this period

Follow-up
- 1 year following the end of intervention or transition to another setting

Baseline at initial assessment or intake
- End of intervention(s) or transition to another setting

Figure 2: Timeline for data collection

Panel 3: How to access the standard set resources
A standard set reference guide, flyer, and data dictionary is available free of charge and can be accessed via ICHOM Connect. The reference guide defines each outcome domain and provides a detailed description of the recommended measures, case-mix factors, and timepoints.

Embed monitoring into the clinical process and reduce the risk of missing data due to drop-out before follow-up. The group recognises that effective session by session measurement requires well established systems and can otherwise be perceived as unnecessarily burdensome.

To mark the end of an active treatment cycle, outcomes should be measured upon transition into adult services, into a different level of care (eg, from outpatient to inpatient care), or upon completion when no further activities are planned (figure 2). Outcomes should then be measured again at a follow-up assessment 1 year after baseline measurements were taken. This can enable important insights into longer-term outcomes, but could require substantial adjustments to the way services are currently organised and funded.

Strengths of the standard set
Of the working group’s 27 voting members, four were young adults aged between 18 and 24 years, and two were parents. This was the first ICHOM working group to involve young people with lived experience of service use—rather than only their parents—as full working group members at all stages of the consensus-building process, including the teleconferences, the adapted Delphi exercise (a method for reaching group consensus), and subsequent surveys. Although limited familiarity with specific methodological questions around outcome measurement can form a barrier to the meaningful participation of lived-experience experts, the teleconferences helped foster a common understanding within the group ahead of each round of voting. The working group chair (MW) solicited input from all call participants on all key discussion points to promote equal participation and manage power imbalances. Additional feedback from a wider group of young people and parents was sought (via the open review process) towards the end of the consensus-building process. Variable requirements for ethical review for this stakeholder group meant that the survey was only accessible in three countries, while the professional survey was accessible globally. This difference in availability led to a comparatively small sample of lived-experience participants (n=24) in the open review process. However, separate analysis and review of feedback from the professional and lived-experience surveys meant that the working group was able to consider each feedback stream in its own right.

The working group included experts from low-income and middle-income countries (LMICs) who have rarely been represented by similar initiatives to date (appendix p 3). Although 90% of children and young people live in LMICs, 90% of research on youth mental health comes from high-income countries. Experts from LMICs were also consulted through the open review process (appendix p 18). A range of local needs, challenges, and cultural factors could thus be considered. The high endorsement of the standard set in the open review process underscores its relevance and acceptability beyond the immediate working group.

Implementation
Resources that can guide the implementation of this standard set are available from ICHOM (panel 3). Members of the working group will form a steering group to oversee the implementation of the standard set and to review recommendations while taking into consideration the lessons from pilot initiatives and new developments in the field. In several health systems (eg, from Australia, Canada, the Netherlands, the UK, or the USA), the routine collection of outcome data is becoming increasingly embedded, which should facilitate the adoption of the standard set but can also cause issues of alignment with existing local or national systems.

As services adopt routine outcome monitoring, it is imperative that data are handled safely and securely in accordance with relevant data protection frameworks, and that informed consent protocols are in place.

This standard set was designed for clinical practice and might be suitable for use in clinical trials, depending on the requirements and the type of intervention tested. Since work on this standard set has ended, the Wellcome Trust and the National Institute of Mental Health (NIMH) have recommend the RCADS-25 as an outcome measure for research with children and adolescents experiencing depression or anxiety (suitable for ages 8–18). The funders suggest that the PHQ-9 and GAD-7 could be used instead of the RCADS-25 with adolescents and young adults, where this is judged developmentally appropriate. The funders recommend the WHODAS 2.0 as a functioning measure for young adults from the age of 18 years. A standard set designed specifically for clinical trials in adolescent depression is under development as part of the International Network for Research Outcomes in Adolescent Depression Studies (IN-ROADS) initiative at
the Hospital for Sick Children (Toronto, ON, Canada). UNICEF is leading another complementary initiative focusing on mental health outcomes for adolescents with anxiety or depression in population surveys. Although there is a unique opportunity to encourage further harmonisation, a degree of divergence might persist because of differences in priorities, processes, and methodologies (eg, outcome measures for clinical trials might not need to be freely available). The present work adds to existing review efforts to identify meaningful, feasible, and acceptable outcome measures suitable for use with children and young people in practice settings.

The steering group will consider opportunities for alignment with complementary initiatives, as well as with existing ICHOM standard sets (eg, with a view to linking to existing review efforts). To promote uptake of this standard set, simplicity and feasibility were prioritised over detail and specificity. Brief and freely available instruments were prioritised over more complex and, at times, more established ones, and short forms over long versions. As the former tend to be less widely validated than the latter, it is hoped that the standard set will accelerate validation efforts and generate new data for their psychometric properties.

To maintain simplicity, the set does not recommend separate measures for different age groups. Although the selected measures have generally been validated in children and young people aged between 6 and 18 years, the set is less specifically tailored to the experiences of those aged between 18 and 24 years. As the set moves into implementation, the steering group will consider whether, on balance, the gains from increased feasibility can be seen to justify this design.

None of the recommended measures are perfect. The goal was to identify the best-possible suite of instruments within the working group’s feasibility and psychometric criteria, based on the evidence available at the time, as a starting point for generating wider insights into how outcome measurement might be strengthened in the future. Although the working group considered sensitivity to change an essential measurement property, its appraisal was hindered by the scarcity of data and objective appraisal guidelines. The International Society for Quality of Life Research (ISOQOL) recommends that for longitudinal research, patient-reported outcome measures “should have evidence of responsiveness, including empirical evidence of changes in scores consistent with predefined hypotheses” about the expected treatment outcome.

Although the working group considered such evidence when available, no standard thresholds could be applied to determine whether sensitivity was sufficient. An important area that could not be considered as part of the appraisal is the measurement invariance of the selected tools across languages and cultural backgrounds, and the extent to which cultural differences might affect the validity of the selected measurement instruments. In the absence of invariance, comparisons between different groups are not fully meaningful, and the working group hopes that this initiative will enhance data availability and spur efforts to examine how consistently the recommended measures track their designated outcome concepts across cultural and language contexts.

The working group encountered challenges with identifying suitable measures of functioning, with common issues including overlap in item coverage between symptom and functioning, a perceived overemphasis on bodily functions as opposed to psychosocial functioning, an absence of validation across the full age span or in relevant clinical populations, and cross-cultural validity. Overall, additional research is needed to understand how suitable the recommended measures are for capturing change over the course of treatment, how suitable these measures are in different real-world clinical settings (eg, primary vs specialist care), and how acceptable to services and service users.

Limitations and areas for future research
The scope of this standard set is limited to anxiety, depression, OCD, and PTSD. The working group has sought to make a minimum recommendation focusing on core outcomes for these target disorders. Clinical judgment is warranted when tracking additional symptoms over time that are not covered by the recommended outcome measures. Three complementary ICHOM standard sets are available that focus on psychotic disorders (covering bipolar disorder), personality disorders, and substance use and addictive behaviour disorders in young people and adults. The present standard set captures the presence of these disorders and other comorbid presenting problems at baseline via the current view screening tool for the purpose of case-mix adjustment.

This standard set was developed by a small working group of consistent members that convened at frequent intervals. There was not complete parity in representation across different strata of experts, including young people and parents. In future efforts, large-scale Delphi surveys might be more suitable for consulting equal numbers of different stakeholder groups, albeit at the expense of more in-depth and continuous group deliberation.

Because of the nature of consensus building, compromises had to be made and not all individual views and priorities could be reflected in the final standard set. For example, two additional outcome domains (ie, coping, and interference of treatment with daily life) reached consensus among the group’s lived experience experts, but not within the wider group. To promote person-centred care, services could consider tracking additional outcomes on the basis of shared decision making with children, young people, and families, either through suitable standardised measures or by including person-alised outcome measures that track progress in relation to individual problems or treatment goals.

To promote uptake of this standard set, simplicity and feasibility were prioritised over detail and specificity, and short forms over long versions. As the former tend to be less widely validated than the latter, it is hoped that the standard set will accelerate validation efforts and generate new data for their psychometric properties.

To maintain simplicity, the set does not recommend separate measures for different age groups. Although the selected measures have generally been validated in children and young people aged between 6 and 18 years, the set is less specifically tailored to the experiences of those aged between 18 and 24 years. As the set moves into implementation, the steering group will consider whether, on balance, the gains from increased feasibility can be seen to justify this design.
Call for action

This standard set provides the first global guideline for promoting the quality and consistency of routine outcome measurement for children and young people with anxiety, depression, OCD, or PTSD. It is person centred and devised specifically for use in clinical practice, with special attention given to acceptability and feasibility, including in resource-poor contexts. It also has great relevance to the provision of mental health care outside the health service system, such as in school settings. It forms an essential step towards enhancing evidence on service effectiveness, enabling comparisons and benchmarking of results across care systems, and promoting care quality in mobilising a comprehensive, forceful, and evidence-based response to the global burden of anxiety and depression. Future research should continue to expand the evidence base in relation to the sensitivity to change and measurement invariance of the included measures, and implementation initiatives should provide feedback on the relevance and acceptability of this recommendation to practitioners and service users. Both will be crucial to ensure that this standard set makes a viable recommendation that meaningfully captures change for children and young people across contexts.

Contributors

KRK, MW, and SC oversaw the study design. KKR and SC led data acquisition through the literature reviews, Delphi surveys, and open review surveys, as well as data analysis and interpretation. KKR took responsibility for the integrity of the data and the accuracy of data analysis. All authors attended teleconferences, and all but the core project team (KKR, MW, SC) completed anonymous votes and feedback surveys as part of the Delphi process and formed consensus on the final recommendation. KKR drafted the manuscript with support from SC, and all co-authors critically reviewed the working draft and agreed to the revisions and final submission. Administrative, technical, and organisational support was provided by SC.

Declaration of interests

AMA receives royalties from Oxford University Press for the Anxiety Disorders Interview Schedule (ADIS), child and parent versions. PB is involved with the development of routine outcome measurement in public mental health across Australia. He chairs the National Child and Adolescent Mental Health Information Development Expert Advisory Group which supports routine outcome measurement and benchmarking between organisations, but has no financial interest in any system. SC was an employee of the International Consortium for Health Outcomes Measurement (ICHOM) during this study. ICHOM received grants from the New South Wales Agency for Clinical Health Outcomes Measurement (ICHOM) during this study. ICHOM received grants from the New South Wales Agency for Clinical Health Outcomes Measurement (ICHOM) during this study. KKR received personal fees that aimed to develop Japanese versions of scales, including the Spence Lundbeck, Actelion, UCB, Otsuka, Grunenthal, ViiV, outside the Almirall, Allergan, Stallergene, Genzyme, Pierre Fabre, AstraZeneca, GlaxoSmithKline, HRA Pharma, Roche, Boehringer Ingelheim, Bayer, fees from Eli Lilly, Bristol Myers Squibb, Servier, sales force, and all co-authors critically reviewed the working draft and agreed to the revisions and final submission. Administrative, technical, and organisational support was provided by SC.

References

Bickman L. A measurement feedback system (MFS) is necessary to improve mental health outcomes. *J Am Acad Child Adolesc Psychiatry* 2008; 47: 1114–19.


Clarke M, Williamson PR. Core outcome sets and systematic reviews. *System Rev* 2016; 5: 11.


© 2020 Elsevier Ltd. All rights reserved.