RUNNING HEAD: MEASURING RESPONSE TO CLINICAL CARE

Accepted for publication by the Lancet Psychiatry

Measuring Response to Clinical Care in Children and Young People with Anxiety, Depression, OCD or PTSD: An International Standard Set of Outcome Measures

Karolin Krause, Sophie Chung, Abiodun O. Adewuya, Anne Marie Albano, Rochelle Babins-Wagner, Laura Birkinshaw, Peter Brann, Cathy Creswell, Kathleen Delaney, Bruno Falissard, Christopher B. Forrest, Jennifer L. Hudson, Shin-ichi Ishikawa, Meghna Khatwani, Christian Kieling, Judi Krause, Kanika Malik, Vania Martínez, Faraz Mughal, Thomas H. Ollendick, Say How Ong, George C. Patton, Ulrike Ravens-Sieberer, Peter Szatmari, Evie Thomas, Lucas Walters, Bryan Young, Yue Zhao, Miranda Wolpert

Author Affiliations

Evidence Based Practice Unit, University College London and Anna Freud National Centre for Children and Families, London, United Kingdom (KK, PhD)

International Consortium for Health Outcomes Measurement (ICHOM), London, United Kingdom (SC, BA)

Lagos State University College of Medicine, Ikeja, Lagos, Nigeria (AOA, Prof, FWACP)

New York State Psychiatric Institute and Columbia University Vagelos College of Physicians and Surgeons, Columbia University, New York, New York, USA (AMA, Prof, PhD)

Calgary Counselling Centre, Calgary, AB, Canada; Faculty of Social Work, University of Calgary, Calgary, Alberta, Canada (RBW, PhD)

Eastern Health Child and Youth Mental Health Service & Monash University, Melbourne, Victoria, Australia (PB, PhD)

Young Minds, London, United Kingdom (LB, HBA)

Departments of Experimental Psychology and Psychiatry, University of Oxford, Oxford, United Kingdom (CC, Prof, PhD)

Rush College of Nursing, Chicago, Illinois, USA (KD, Prof, PhD)

Université Paris-Saclay, Université de Versailles Saint-Quentin-en-Yvelines, Institut national de la santé et de la recherche médicale (INSERM), Centre de recherche en Epidémiologie et Santé des Populations (CESP), Villejuif, Île-de-France, France (BF, MD & PhD)

Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA (CBF, Prof, PhD)

Centre for Emotional Health, Department of Psychology, Macquarie University, Sydney, New South Wales, Australia (JLH, Professor, PhD)

Faculty of Psychology, Doshisha University, Kyoto, Japan (SI, Prof, PhD)

Department of Psychology, University of Delhi, New Delhi, Delhi, India (MK, BA)

Hospital de Clínicas de Porto Alegre, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil (CK, MD)

Child and Youth Mental Health Service, Children's Health Queensland Hospital and Health Service, Brisbane, Queensland, Australia (JK, MA)

Sangath, New Delhi, Delhi, India (KM, PhD)

Centro de Medicina Reproductiva y Desarrollo Integral del Adolescente (CEMERA), Facultad de Medicina, Universidad de Chile; Millennium Nucleus to Improve the Mental Health of Adolescents and Youths (Imhay); Millennium Institute for Research in Depression and Personality (MIDAP), Santiago de Chile, Región Metropolitana, Chile (VM, Prof, PhD)

School of Primary, Community and Social Care, Keele University, Keele ,Staffordshire, United Kingdom (FM, MBChB)

Child Study Center, Department of Psychology, Virginia Tech, Blacksburg, Virginia, USA (THO, Prof, PhD)

Department of Child and Adolescent Psychiatry, Institute of Mental Health, Singapore, Singapore (SHO, MMED)

Eastern Health Child Youth Mental Health Service, Victoria, Australia (ET, BA)

Murdoch Children's Research Institute, Melbourne; Department of Paediatrics, The University of Melbourne, Melbourne, Victoria, Australia (GCP, Prof, MD)

Department of Child and Adolescent Psychiatry, Psychotherapy, and Psychosomatics, University Medical Center Hamburg-Eppendorf, Hamburg, Germany (URS, Prof, PhD)

The Hospital for Sick Children, the University of Toronto; the Cundill Centre for Child and Youth Depression, Centre for Addiction and Mental Health, Toronto, Ontario, Canada (PS, Prof, MD)

Jack.org, Ottawa, Ontario, Canada (LW, B.Sc.; BY)

Teaching and Learning Evaluation and Measurement Unit, The University of Hong Kong, Hong Kong (YZ, EdD)

Wellcome Trust, London, Greater London, United Kingdom (MW, Prof, DClinPsy)

Corresponding Author: Karolin Krause, Evidence Based Practice Unit, Anna Freud National Centre for Children and Families, The Kantor Centre of Excellence, 4-8 Rodney Street, London N1 9JH, United Kingdom; karolin.krause.16@ucl.ac.uk, +44 7850 296780.

Contributors

KK, MW, and SC oversaw the study design. KK and SC led data acquisition through the literature reviews, Delphi surveys, and Open Review surveys, as well as data analysis and interpretation. KK takes responsibility for the integrity of the data and the accuracy of data analysis. All authors attended teleconferences, and all but the core project team (KK, MW, SC) completed anonymous votes and feedback surveys as part of the Delphi process and formed consensus on the final recommendation. KK 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 across Australia of routine outcome measurement in public mental health. He chairs the National Mental Health Child and Adolescent Information Development Expert Advisory Group. There is an interest in supporting routine outcome measurement and benchmarking between organisations. This does not influence the content of the submitted work. SC was an employee of the International Consortium for Health Outcomes Measurement (ICHOM) during the conduct of this study. ICHOM received grants from the Government of New South Wales Agency for Clinical Innovation, Australia; Providence Health & Services, USA; Västra Götaland Regional Council, Sweden; and National Health Service, England, United Kingdom, in support of this work. CC has conducted several studies and ongoing evaluations of commonly used scales, including the Spence Children's Anxiety Scale (SCAS), Revised Children's Anxiety and Depression Scale (RCADS) and the Child Anxiety Interference Scale (CAIS). BF reports personal fees from E. Lilly, BMS, Servier, SANOFI, GSK, HRA, Roche, Boeringer Ingelheim, Bayer, Almirall, Allergan, Stallergene, Genzyme, Pierre Fabre, AstraZeneca, Novartis, Janssen, Astellas, Biotronik, Daiichi- Sankyo, Gilead, MSD, Lundbeck, Stallergene, Actelion, UCB, Otsuka, Grunenthal, ViiV, outside the submitted work. JLH is one of the developers of the Child Anxiety Life Interference Scale (CALIS), which is a freely available measure and there is no financial conflict of interest. SI has participated in several studies which aimed to develop Japanese versions of scales, including the Spence Children's Anxiety Scale, Children's Depression Scale; and the Social Phobia and Anxiety Inventory for Children. KK received personal fees from the International Consortium for Health Outcomes Measurement (ICHOM) during the conduct of the study. URS is one of the developers of the KIDSCREEN-10, which is a freely available measure and there is no financial conflict of interest. Since May 2019, MW is head of the new Mental Health Priority Area at the Wellcome Trust, which may be developing Standard Sets in mental health in the future. She has been involved in the development of the Current View Tool, which is a freely available measure and there is no financial conflict of interest. MW was previously Head of the Child Outcomes Research Consortium (CORC), which advises on measurement in child mental health; and an advisor to NHS England on informatics. AOA, RBW, LB, KD, CBF, MK, CK, JK, FM, KM, VM, THO, SHO, ET, GCP, PS, LW, BY, YZ have nothing to disclose.

Acknowledgments

The authors would like to gratefully acknowledge all individuals and organisations who took the time to share or complete the Open Review feedback survey towards improving the Standard Set recommendation. The authors would also like to thank Luz Sousa Fialho for providing administrative support, Katherine G. Young for providing research support, and Christiana Ilesanmi for providing both administrative and research support to this project, as well as Zofia Das-Gupta for providing guidance on the application and management of the ICHOM process until May 2019. The authors would like to thank one further service user representative who participated in the early stages of the process, including the Delphi process, but did not wish to be a named co-author.

Funding

This work has been funded by the National Health Service (NHS) England; Västra Götaland Regional Council, Sweden; the Government of New South Wales Agency for Clinical Innovation, Australia; and Providence Health & Services, USA. The funders had no role in the study design, data analysis, interpretation, or writing of the report. The funders did support data collection by sharing the Open Review survey for practitioners within their networks.

CC was supported by a National Institute for Health Research (NIHR) Research Professorship (RP_2014-04-018). FM was supported by an NIHR In-Practice Fellowship (IPF-2017-11-002), and is supported by a NIHR School for Primary Care Research GP Career Progression Fellowship. The views expressed in this manuscript of those of the authors and not necessarily those of the NHS, NIHR, or Department for Health and Social Care. VM received funding by the Millennium Science Initiative of the Ministry of Economy, Development and Tourism, grant Millennium Nucleus to Improve the Mental Health of Adolescents and Youths, Imhay, and the Innovation Fund for Competitiveness [Fondo de Innovación para la Competitividad, FIC], Project IS130005. CK is a Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) researcher and an Academy of Medical Sciences Newton Advanced Fellow.

Keywords

Anxiety; Depression; Stress Disorders, Post-Traumatic; Obsessive-Compulsive Disorder; Patient Reported Outcomes; Outcome Measures; Patient-Relevant Outcome; Quality Improvement; Reference Standards; Delphi Technique; Child Health; Adolescent Health; Core Outcome Set.

Wordcount

4760 words

Abstract

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 (ICHOM) has already developed a consensus-based Standard Set of outcomes for anxiety and depression in adults (including the PHQ-9, GAD-7 and WHODAS 2.0). This paper reports on recommendations specifically for anxiety, depression, obsessive-compulsive disorder (OCD), and post-traumatic stress disorder (PTSD) in children and young people (CYP) aged 6-24 years. An international ICHOM working group of 27 clinical, research, and lived-experience experts formed consensus through teleconferences, a Delphi exercise, and iterative anonymous voting. A systematic review identified 70 possible outcomes and 107 relevant measurement instruments. Measures were appraised for their feasibility in routine practice (i.e., brevity, free availability, validation in CYP, translation) and psychometric performance (i.e., validity, reliability, 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 RCADS-25, OCI-CV, CRIES-8/13, C-SSRS, KIDSCREEN-10, CGAS, and CALIS. 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.

Measuring Response to Clinical Care in Children and Young People with Anxiety, Depression, OCD or PTSD: An International Standard Set of Outcome 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 (1). 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 (2–4). Despite an increase in mental health care provision over recent decades, service systems have failed to reduce the prevalence of these disorders in children and young people (CYP; 5). The global response 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 (6,7). Valid data on treatment outcome is an essential facet in evaluating care effectiveness and can inform the setting of strategic targets for health systems, comparisons between systems and services, and clinical decision-making on a case-by-case basis (8,9).

Currently there is neither agreement, nor global guidance on how best to track the response to clinical care for anxiety and depression in CYP. Uptake of routine outcome measurement remains low, and where it occurs, there is considerable variation in outcomes, instruments, and assessment time points, with a recent review recording 20 different measures used to assess primary outcome across 38 studies of routine treatment for youth anxiety and depression (10). 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 (11). 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 service user representatives, practitioners, and researchers to build consensus on a set of outcome domains, measurement instruments, case-mix factors (i.e., 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 to recommend.

Panel 1. The ICHOM Approach to Standard Set Development

ICHOM is a non-profit organisation specialised in the development of condition-specific Standard Sets for clinical practice. ICHOM has supported the development of more than 28 existing Standard Sets, including one for adult depression and anxiety (12). Outcome measurement is approached within a framework of person-centred and value-based healthcare, where value is defined as the health outcomes achieved, relative to the resources invested, rather than the volume of services delivered (13). 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 service user needs and treatment options are at least partly shaped by the principal presenting problems (13,14). 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; balance a comprehensive approach to tracking outcome with a feasible recommendation that services can reliably implement; and be responsive to quality improvement efforts.

An existing ICHOM Standard Set for adult anxiety and depression covers young people from the age of 14 years (12; Appendix pg. 3). It includes a recommendation for tracking symptom change via the Generalized Anxiety Disorder 7-item Scale (GAD-7; 15), and the depression subscale of the Patient Health Questionnaire (PHQ-9; 16); as well as functioning through the World Health Organization Disability Schedule 2.0 (WHODAS 2.0) 12-item short form (17; the full recommendation is available from ichom.org). The present Standard Set aims to complement this effort by providing a set of recommendations specifically tailored for use with CYP. The combination of the two Sets will provide for transition from youth into adult services at any point between the ages of 14 and 24 years, allowing for local variation in transition and judgements about which Set is most suitable for different ages.

This Standard Set is designed for CYP aged 6 to 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 levels of negative affect, with OCD and PTSD long classified as anxiety disorders (18). It 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-, parent-, and clinician-reported outcome measures, to account for the different perspectives these reporters tend to provide (19-21). Parent-report also serves as a proxy where CYP are unable to complete measures due to young age or developmental constraints (21). A more detailed discussion of the scope of this Standard Set is provided in the Appendix (pg. 3).

Panel 2. Methodology

The Working Group comprised 27 experts from 13 countries, including service users, parents or carers (henceforth "parents"), mental health practitioners and researchers working within relevant disciplines (e.g., psychiatry, epidemiology, psychometrics; see Appendix pg. 3-4). A central project team (KK, SC, MW) coordinated and facilitated the consensus-building process and completed supporting research tasks, but did not vote on the consensus recommendation. The Appendix (pg. 3-18) provides a detailed description of the process and methodology. A flow chart is provided in figure 1.

General process. Over 14 months, the Working Group completed a structured and evidence-informed consensus-building process (figure 1). The group convened for eight teleconferences, completed a three-round Delphi exercise to select outcome domains (22,23), and participated in iterative rounds of anonymous voting to arrive at recommendations for outcomes, measurement instruments, case-mix factors, and time points (Appendix pg. 4-5). The process was informed by sequential research inputs performed by the central project team (Appendix pg. 5-18).

Selection of outcome domains. In line with methodological recommendations by the Core Outcome Measures in Effectiveness Trials (COMET) initiative (24), 70 possible outcomes and 507 measurement instruments were identified through multiple avenues, including a systematic review of 257 treatment outcome studies, a narrative review of supplemental sources (e.g., cohort studies; qualitative outcome research, instrument banks), and break-out groups with service user representatives (Appendix pg. 5-8)

Selection of measurement instruments. Once the Group reached consensus on outcomes for inclusion, 107 thematically relevant measures were systematically appraised to identify those most suitable for tracking the selected outcomes over time. Appraisal criteria included *relevance*; (i.e., comprehensive coverage of the selected outcome domain); *feasibility* and *acceptability* (i.e., completion within less than 20 minutes; free availability for use in clinical settings, including in paper-and-pencil format; prior validation in CYP; translation into at least a second language); and *psychometric performance* (i.e., interrater or retest reliability, and internal consistency above 0.70; evidence of sensitivity to change), in line with International Society for Quality of Life Research (ISOQOL) recommendations (25; Appendix pg 8-16). The instruments judged to best satisfy these criteria were shortlisted, and their relative 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 time points. Based on the systematic review and existing ICHOM Standard Sets, the central project team compiled a list of possible case mix factors, and conducted a rapid review

of reviews examining predictors, mediators, and moderators of treatment response (Appendix pg. 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 time points, 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 policymakers from 45 countries, as well as 24 young people and parents from Denmark, the United States (US), and the United Kingdom (UK; see Appendix pg. 18).

Recommendations of the Working Group

Outcomes and Measures

As per Working Group consensus, the Standard Set recommends tracking response to treatment across the three outcome domains of *symptoms*, *suicidal thoughts and behaviour*, and *functioning*, using seven primarily self-reported instruments. These were selected for fulfilling most or all of the appraisal criteria (table 1). A detailed discussion of instrument properties, performance against appraisal criteria, and accessibility is provided in the Appendix (pg. 19-22). During the Open Review, 75% of the 463 practitioners and researchers who provided feedback stated their overall confidence in the recommended outcomes and instruments and participants with lived experience confirmed the importance of outcomes included (75-100% of participants rated each included outcome domain as important) and the acceptability of the recommended measures (100% of 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.

Insert Table 1

Symptoms. The Set recommends measuring anxiety and depression symptoms through youth and parent-report for all CYP in scope. 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 Set recommends the 25-item short form of the Revised Children's Anxiety and Depression Scale (RCADS-25; 26) for youth and parent-report. Based on the appraisal criteria, the group initially selected the scale's 47-item version (27), but decided to recommend the short form to reduce respondent burden. The RCADS-25 consists of 15 items tracking anxiety symptoms, and 10 items tracking 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 (28).

Symptoms of OCD and PTSD should be tracked separately for CYP 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 Compulsive Inventory for Children (OCI-CV; 29); and the Children's Revised Impact of Events Scale (CRIES; 30) for youth (CRIES-8) and parent report (CRIES-13). Both measures have been applied in Europe, Asia, and the Americas. A parent version of the OCI-CV is not currently available.

It is important to note that the above-mentioned symptom measures are recommended for the purpose of tracking treatment outcomes over time. They are not considered primarily for the purpose of diagnosing presenting problems, and are not intended to replace a thorough clinical assessment using state-of-the-art diagnostic tools. The latter require additional properties related to diagnostic validity (e.g., sensitivity and specificity), which the Working Group did not explicitly consider during measure appraisal.

Suicidal thoughts and behaviour. Consensus was reached on measuring suicidal thoughts and behaviour in all young people aged 10 years and older (unless considered inappropriate) using the Columbia Suicide Severity Rating Scale (C-SSRS) Recent Self-Report Screener (31) – a short self-report version of the clinician-administered C-SSRS interview protocol. The measure consists of six items tracking the severity of suicidal ideation and behaviour in the previous month. The self-report screener has not been validated in CYP, but the clinician-rated C-SSRS has demonstrated good internal consistency, interrater reliability, and sensitivity to change in adolescent samples (31).

Functioning. Functioning describes a child's ability to engage in typical activities and meet role demands in line with age-specific socio-cultural norms (32,33). None of the identified functioning measures fully satisfied the Working Group's requirements, and consensus was reached on mitigating this 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, while condition-specific measures may 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 (i.e., item 8: *I have trouble sleeping*; and item 9, *I feel scared if I have to sleep on my own*).

The KIDSCREEN-10 was selected as a generic measure of global functioning, to be completed by CYP and parents. This unidimensional 10-item index of HRQoL tracks functioning in relation to physical health and energy levels, leisure activities, social and family relationships, and cognition (34). The more comprehensive KIDSCREEN-52 was originally developed through a process of cross-cultural harmonisation involving 13 European countries. Its 10-item short form has been applied in Asia, Eastern Africa, Europe, North and Latin America. Although originally designed for epidemiological studies, the KIDSCREEN-10 has been shown to discriminate well between CYP with high and low levels of functioning, with few ceiling or floor effects (34). In addition to the KIDSCREEN-10, the Children's

Global Assessment Scale (CGAS) was selected as a brief clinician-rated measure of global functioning (35). On this widely used measure, clinicians preform a single rating by locating CYP on a scale from 1 to 100, placing them into one of ten categories, from 1 – 10 (*extremely impaired*) to 91 – 100 (*doing very well*).

The Children's Anxiety Life Interference Scale (CALIS; 36) was selected to measure condition-specific impairment via 9 items (10 for the parent version) that describe instances of anxiety impacting on functioning at home, at school, on social life, or on activities. The CALIS currently only covers anxiety-related impairment, and the Working Group did not identify any eligible measure that tracks depression-specific impairment, or captures impairment from both anxiety and depression. However, the CALIS author team recently revised the measure to cover impairment from anxiety *and* depression, and a validation study involving children, young people, and parents in community, school, and clinical settings is ongoing. As soon as validation results 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.

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 Working Group consensus, services should record demographic, clinical, complexity, and intervention factors, through a mix of self-, parent-, and clinician-report (table 2). Beyond the Working Group, the suggested case-mix factors were endorsed by more than 80% of practitioners and researchers who provided feedback during the Open Review. These factors represent a minimum that should be assessed by all those providing relevant care, and services may wish to add other indicators to meet local information needs.

Insert Table 2

Demographic factors. The Set recommends recording age, gender, ethnic minority status, socio-economics status, and the child's living situation. Services should record the sex assigned at birth, and the gender reported by CYP. Socio-economic status should be measured by recording the highest level of education completed by any of the CYP's parents, as a widely accepted proxy that can be mapped onto the International Standard Classification of Education for international comparisons (ISCED; 37). The Set includes one question about CYP's living situation, and two questions capturing ethnic minority and marginalised group status via self-report (table 2).

Clinical factors. Several studies suggest that symptom burden, symptom duration, and the presence of comorbidities affect treatment response in CYP with depression or anxiety (38–47; Appendix pg. 17-18). The group recommends recording principal and comorbid presenting problems by administering the 30 problem descriptions of the Current View Tool (48). While not equivalent with formal diagnoses, these broadly align with the diagnostic categories of the

ICD-11 for pediatric populations (49). ICHOM standards are available for recording symptom duration and prior service use (table 2).

Complexity factors. Research suggests that parental mental health influences treatment response in CYP with anxiety, and depression (43,50-52; Appendix pg. 17-18). The Set includes two questions about experiences with or diagnoses of mental health problems in the immediate family, and prior use of mental health services by the reporting parent. While evidence is limited and conflicting on the influence of adverse experiences (e.g., childhood maltreatment) on treatment response (43,53), the Set recommends recording trauma history via the *Selected Complexity Factors* section of 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 intervention focus (i.e., in terms of who is actively involved), treatment modality and prescribed medication (i.e., as per lists of options compiled by the Working Group and through Open Review feedback), and intervention setting (i.e., in terms of whether or not treatment was delivered via a digital platform or inpatient care, as opposed to other settings). Services may wish to record additional detail on intervention characteristics to meet local, regional, or national information needs.

Measurement Time Points

Timelines for measurement are highly practical considerations, 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 local or national data collection. The suggested timepoints were widely endorsed by over 80% of practitioners and researchers participating in the Open Review.

As per Working Group consensus, all case-mix factors and outcomes should be measured at assessment or intake (i.e., baseline), or as near to these time points as possible where services wish to collect data at second contact (figure 2). As a guideline, all outcome measures should be administered every three months following baseline. Services are encouraged to consider more frequent intervals, including session-by-session measurement, to help embed monitoring into the clinical process (54), and reduce the risk of missing data due to drop-out prior to follow-up. The group recognises that effective session-by-session measurement requires well established systems and can otherwise be perceived as unduly 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 (e.g., 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, one year after baseline. This

can enable important insights into longer-term outcomes, but may require significant adjustments to the way services are currently organised and funded.

Insert Figure 2.

Strengths of the Standard Set

Of the Working Group's 27 voting members, four were young adults between the ages of 18 and 24, and two were parents. This was the first ICHOM Working Group to involve young people with lived experience of service use – rather than just their parents – as full Working Group members at all stages of the consensus-building process, including the teleconferences, Delphi exercise, and subsequent surveys. While limited familiarity with outcome measurement has been described as a barrier to the meaningful participation of lived-experience experts (55), 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 towards the end of the consensus-building process, through the Open Review. 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 led to a comparatively small sample of Open Review participants with lived experience (N = 24). 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- and middle-income countries (LMIC, Appendix pg. 3), who have rarely been represented by similar initiatives to date (56). While 90% of CYP live in LMICs, 90% of research on youth mental health currently comes from high-income countries (57). Experts from LMIC were also consulted through the Open Review (Appendix pg. 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 underscores its relevance and acceptability beyond the immediate Working Group.

Implementation

Working Group members will form a steering group to oversee the Standard Set's implementation, and to review recommendations in light of learning from pilot initiatives and new developments in the field. In health systems such as Australia, Canada, the Netherlands, the UK, or the US, the routine collection of outcome data is becoming increasingly embedded (58,59), which should facilitate the adoption of the Standard Set, but can also cause issues of alignment with existing local or national systems (e.g., 60,61). As services adopt routine outcome monitoring, it is imperative that data is 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 may or may not 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, along with the PHQ-9, GAD-7 and WHODAS for older youth. A Standard Set specifically for clinicial trials for adolescent depression is currently under development at the Hospital for Sick Children in Toronto (62). The United Nations Children's Fund (UNICEF; 63) is leading another complementary initiative focusing on measuring health outcomes for adolescents with anxiety or depression in population surveys. While there is a unique opportunity to encourage further harmonisation, a certain degree of divergence may persist in light of different priorities, processes, and methodologies (e.g., outcome measures for clinical trials do not need to be freely available). The present work further adds to existing review efforts to identify meaningful, feasible, and acceptable outcome measures suitable for use with CYP in practice settings (64-68). The steering group will consider opportunities for alignment with complementary initiatives, as well as with existing ICHOM Standard Sets, for example with a view to linking scores obtained from the CYP and adult Sets for the purpose of longitudinal analysis (e.g., 69).

Panel 3. How to Access the Standard Set Resources

A Standard Set reference guide, flyer, and data dictionary are available free of charge and can be accessed via ICHOM Connect (www.connect.ichom.org). The reference guide defines each outcome domain, and describes the recommended measures, case-mix factors, and time points.

Limitations and Areas for Future Research

The scope of this Standard Set is limited to anxiety, depression, OCD, and PTSD, and the Working Group has sought to make a *minimum* recommendation focusing on core outcomes for these target disorders. Clinical judgement is warranted in tracking additional symptoms over time that are not covered by the recommended outcome measures. Two complementary ICHOM Standard Sets are currently under development that will 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 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 relatively small Working Group of consistent membership, which convened at frequent intervals. There was not complete parity in representation across different strata of experts, including CYP and parents. In future efforts, large-scale Delphi-surveys may be more suitable for consulting equal numbers of different stakeholder groups, albeit at the expense of more in-depth and continuous group deliberation.

As per 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 (i.e., *coping* and *interference of treatment with daily life*) reached consensus amongst the group's service user representatives, but not within the wider group. To promote person-centred care, services may want to consider tracking additional outcomes based on shared decision-making with CYP and families, either through suitable standardised measures, or by including personalised outcome measures that track progress in relation to idiosyncratic presenting problems or treatment goals (70).

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 on their psychometric properties (e.g., 71). To maintain simplicity, the Set does not recommend separate measures for different age groups. While the selected measures have generally been validated in CYP aged 6-18 years, the Set is less specifically tailored to the experiences of 18-24-year-olds. 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. While the Working Group considered sensitivity to change an essential measurement property to consider, its appraisal was limited by a general lack of data and objective appraisal guidelines. 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 (25). While the Working Group considered such evidence where 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 may impact on the validity of the selected measurement instruments (72). In the absence of invariance, comparisons between different groups are not fully meaningful, and the 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.

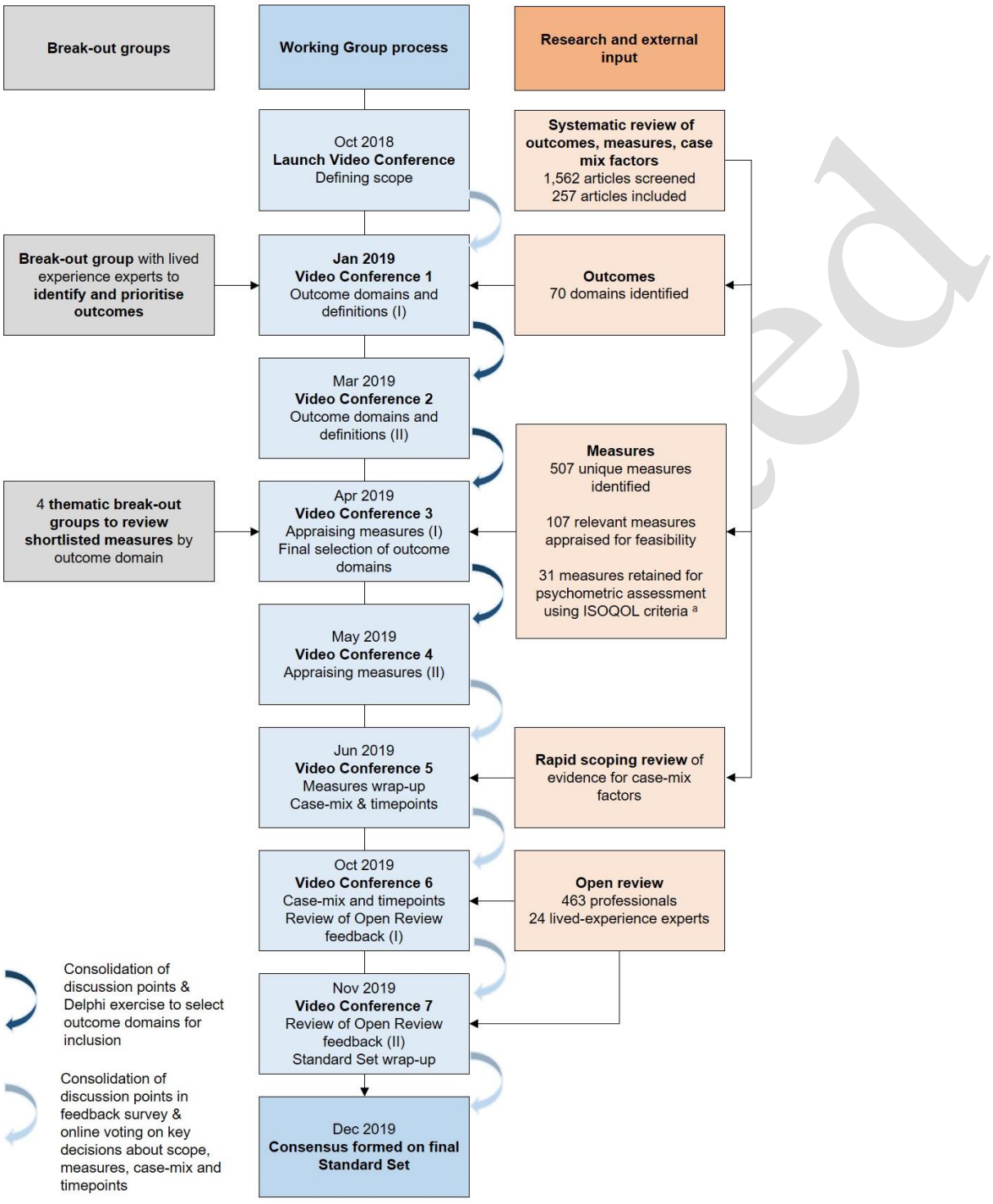
The Working Group encountered challenges with identifying suitable measures of functioning, with common issues including overlap in item coverage between symptoms and functioning, a perceived overemphasis on bodily functions as opposed to psychosocial functioning, lack of validation across the full age span or in 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, and in different real-world clinical settings (e.g. primary versus specialist care), and how acceptable the full question battery is to services and service users.

Call for Action

This Standard Set provides the first global guideline for promoting the quality and consistency of routine outcome measurement for CYP 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 from 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 vital to ensure that this Standard Set makes a viable recommendation that meaningfully captures change for CYP across contexts.

Figure 1. Working Group Process



[a] Shortlisted measures were reviewed against the minimum standards for patient-reported outcome measures recommended by the International Society for Quality of Life Research (ISOQOL; 25).

Table 1. Overview of Recommended Outcome Measures and their Evaluation Criteria

		Relevance	**Feasibility & Acceptability**				**Psychometric Performance** [a]		
		Domain coverage	**Short**	**Free**	**Validated**	**Translated**	**Reliability**	**Validity**	**Sensitivity to change**
		Satisfactory domain coverage at subscale or item level	< 60 items / < 20 min	No licencing costs / restrictions	At least one validation in a CYP sample	>1 language version available	Test-retest reliability (TRT) or inter-rater reliability (IRR) ≥ 0.70	Internal consistency (e.g. Cronbach Alpha ≥ 0.70	Any evidence of sensitivity to change
Outcome	Measure								
Symptoms									
Anxiety & depression	RCADS-25	GAD; MDD; OCD; PD; SAD; SP	25 items (< 10 min)	Yes	Ages 6-18 cl. & non-cl.	4 /16 [b]	TRT (71)	Yes (26,73)	No evidence
OCD	OCI-CV	Doubting/checking; obsessing; hoarding; washing; ordering; neutralising	21 items (< 10 min)	Yes	Ages 6-18 cl. & non-cl.	3+	TRT (29,74–76)	Yes (29,74–78)	Some evidence (29)
PTSD	CRIES-8	Intrusion; avoidance (hyperarousal)	8 /13 items (< 5 min)	Yes	Ages 7-18 cl. & non-cl.	27+	TRT (79)	Yes (79–81)	Some evidence (82) [c]
Suicidal thoughts and behaviour	C-SSRS	Severity of ideation; behaviour (attempts)	3 or 6 items [d] (< 5 min)	Yes	Ages 12-18 b cl. & non-cl	100+	IRR (83) [e]	Yes (31) [e]	Some evidence (31) [e]
Functioning									
Global	KIDSCREEN-10	Physical activity & energy; emotions; leisure time and participation; relationships with parents & peers; cognition; school	10 items (< 5 min)	Yes	Ages 8-18 non-cl.	22+	TRT (34,84)	Yes (34,84,85)	No evidence
Global	CGAS	Global functioning	1 item (< 5 min)	Yes	Ages 4-18	2+	IRR (34,86–88) TRT (35,87)	N.A.	Some evidence (89)
Impact	CALIS	Enjoyable activities; relationships with siblings, parents, friends, peers; sports; schoolwork; distress	9/10 items (< 5 min)	Yes	Ages 6-17 cl. & non-cl.	7	TRT (36)	Yes (36,90,91)	Some evidence (36)

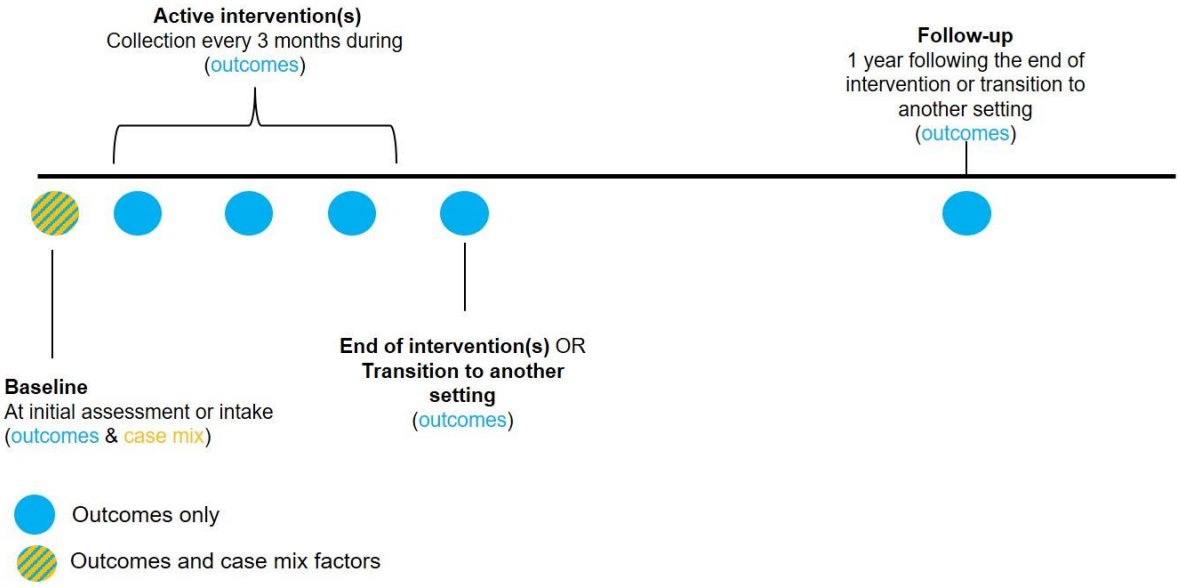
Note. GAD = Generalized Anxiety Disorder; MDD = Major Depressive Disorder (MDD); OCD = Obsessive Compulsive Disorder; PD = Panic disorder; SAD = Separation Anxiety Disorder; SP = Social Phobia. [a] These thresholds are based on the ISOQOL minimum standards for patient-reported outcome measures (25). [b] The RCADS-25 is currently available in four languages on the source website (see Table S12, Appendix pg. 22), but the 47-item long form is available in 16 languages, so that item-level translations are available in more than four languages for those items included in the RCADS-25. [b] These validation studies tested a parent-report version only. [c] These validation studies tested a parent-report version only. [d] Two initial questions serve as screeners, with the three remaining severity items administered only to those endorsing the first two; all young people are asked about suicidal behaviour. [e] To the authors' knowledge, the C-SSRS self-report screener had not been subject to a validation study in CYP at the time of writing. For the psychometric appraisal, studies assessing the psychometric properties of the severity subscale in the clinician-led C-SSRS semi-structured interview schedule were considered instead.

Table 2. Case-Mix Factors in the Standard Set for Pediatric Anxiety, Depression, OCD, and PTSD

	Case-mix factor	Measure	Reporter
Demographic factors	Age [a]	Year of birth	CYP / parent
	Sex [a]	Sex assigned at birth	CYP / parent
	Gender Identity	"Do you think of yourself as...?"	CYP
	Parent or carer education [a]	Highest level of education completed by any of the CYP's parents or carers (ISCED Standards)	CYP / parent
	Ethnicity	Do you consider yourself to be in an ethnic minority where you live?	CYP / parent
	Marginalised group status	Do you consider yourself to be a member of a marginalised group where you live?	CYP / parent
	Living situation	Which of the following people live with you [your child] at your [their] home?	CYP / parent
Clinical factors	Diagnoses and co-morbidities	Measured via the provisional Problems' list of the Current View Tool	Clinician
	Duration of symptoms [a]	For how many months have you [your child] been experiencing [specific condition] symptoms?	CYP / parent
	Prior service use	During the last year, did you [your child] receive any of the following treatments for [specific condition]? Complete separately for (a) medication, (b) psychotherapy, (c) other.	CYP / parent
Complexity factors	Trauma history	Measured via the 'Selected Complexity Factors' of the Current View Tool	Clinician
	Parental mental health	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, substance abuse (for example, alcohol or drugs), schizophrenic disorder, personality disorder, somatoform disorder (unexplained physical symptoms), other)?	Parent
	Parental service use	Have you ever sought help for your mental health?	Parent
	Education/work difficulties	Measured via the 'Contextual Problems' of the Current View Tool	Clinician
Intervention variables	Intervention focus (i)	Who is actively involved in the intervention? Select all that apply.	Clinician
	Intervention focus (ii)	Is the intervention delivered to an individual child / family, or to a group of children / families?	Clinician
	Intervention approach	What is the treatment approach? (Select all that apply)	Clinician
	Prescribed medication	What type of medication is prescribed? (Select all that apply)	Clinician
	Intervention setting (i)	Does this intervention involve an overnight stay at an institution providing mental health support?	Clinician
	Intervention setting (ii)	Does this intervention involve the use of a digital platform?	Clinician

[a] Variables defined and operationalised as per an existing ICHOM standard. More detailed descriptions of each variable, as well as response options can be consulted in the reference guide for the Standard Set, at https://www.ichom.org/portfolio/anxiety-depression-ocd-and-ptsd-in-children-and-young-people/.

Figure 2. Timeline for data collection



References

1. World Health Organization. Depression and Other Common Mental Disorders: Global Health Estimates [Internet]. Geneva; 2017. Available from: https://apps.who.int/iris/bitstream/handle/10665/254610/WHO-MSD-MER-2017.2-eng.pdf [10/01/2020].

2. Kessler RC, Amminger GP, Aguilar-Gaxiola S, Alonso J, Lee S, Üstün TB. Age of onset of mental disorders: A review of recent literature. Curr Opin Psychiatry. 2007;20(4):359–64.

3. Albano AM, Chorpita BF, Barlow DH. Childhood anxiety disorders. In: Mash EJ, Barkley RA, editors. Child psychopathology. 2nd ed. New York, NY, US: The Guilford Press; 2003. p. 279–329.

4. Clayborne ZM, Varin M, Colman I. Systematic Review and Meta-Analysis: Adolescent Depression and Long-Term Psychosocial Outcomes. J Am Acad Child Adolesc Psychiatry. 2019;58(1):72–9.

5. Jorm AF, Patten SB, Brugha TS, Mojtabai R. Has increased provision of treatment reduced the prevalence of common mental disorders? Review of the evidence from four countries. World Psychiatry. 2017;16(1):90–9.

6. Boswell JF, Kraus DR, Miller SD, Lambert MJ. Implementing routine outcome monitoring in clinical practice: Benefits, challenges, and solutions. Psychother Res. 2015;25(1):6–19.

7. Scott K, Lewis CC. Using measurement-based care to enhance any treatment. Cogn Behav Pract. 2015 22(1):49–59.

8. Barkham M, Mellor-Clark J. Bridging Evidence-Based Practice and Practice-Based Evidence: Psychological Therapies. Clin Psychol Psychother. 2003;10:319–27.

9. Bickman L. A Measurement Feedback System (MFS) Is Necessary to Improve Mental Health Outcomes. J Am Acad Child Adolesc Psychiatry. 2008;47(10):1114–9.

10. Bear HA, Edbrooke-Childs J, Norton S, Krause KR, Wolpert M. Systematic Review and Meta-analysis: Outcomes of Routine Specialist Mental Health Care for Young People With Depression and/or Anxiety [published online ahead of print, 2019 Dec 24]. J Am Acad Child Adolesc Psychiatry. 2019;S0890-8567(19)32234-8.

11. Clarke M, Williamson PR. Core outcome sets and systematic reviews. Syst Rev. 2016;5(1):11.

12. Obbarius A, van Maasakkers L, Baer L, Clark DM, Crocker AG, de Beurs E, et al. Standardization of health outcomes assessment for depression and anxiety: recommendations from the ICHOM Depression and Anxiety Working Group. Qual Life Res. 2017;26(12):3211–25.

13. Porter ME. What Is Value in Health Care? N Engl J Med. 2010;363(26):2477–81.

14. Kazdin AE. The Meanings and Measurement of Clinical Significance. J Consult Clin Psychol. 1999;67(3):332–9.

15. 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.

16. Spitzer RL, Kroenke K, Williams JBW. Validation and utility of a self-report version of PRIME-MD: The PHQ Primary Care Study. J Am Med Assoc. 1999;282(18):1737–44.

17. Üstün T, Chatterji S, Kostanjsek N, Rehm J, Kennedy C, Epping-Jordan J, et al. Developing the World Health Organization disability assessment schedule 2.0. Bull World Health Organ. 2010;88(11):815–23.

18. Regier DA, Kuhl EA, Kupfer DJ. The DSM-5: Classification and criteria changes. World Psychiatry. 2013;12(2):92–8.

19. De Los Reyes A, Augenstein TM, Wang M, Thomas SA, Drabick DAG, Burgers DE, et al. The validity of the multi-informant approach to assessing child and adolescent mental health. Psychol Bull. 2015;141(4):858–900.

20. Weisz JR, Kuppens S, Ng MY, Eckshtain D, Ugueto AM, Vaughn-Coaxum R, et al. What five decades of research tells us about the effects of youth psychological therapy: A multilevel meta-analysis and implications for science and practice. Vol. 72, American Psychologist. American Psychological Association; 2017. p. 79–117.

21. Creswell C, Waite P, Hudson J. Practitioner Review: Anxiety disorders in children and young people – assessment and treatment. J Child Psychol Psychiatry Allied Discip. 2020; 61(6): 628-643.

22. Linstone HA, Turoff M. The Delphi Method: Techniques and Applications. Reading, MA: Adison Wesley; 1975.

23. Prinsen CAC, Vohra S, Rose MR, Boers M, Tugwell P, Clarke M, et al. How to select outcome measurement instruments for outcomes included in a "Core Outcome Set" - a practical guideline. Trials. 2016;17(1):1–10.

24. Williamson PR, Altman DG, Bagley H, Barnes KL, Blazeby JM, Brookes ST, et al. The COMET Handbook: Version 1.0. Trials. 2017;18(Suppl 3):1–50.

25. Reeve BB, Wyrwich KW, Wu AW, Velikova G, Terwee CB, Snyder CF, et al. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and

comparative effectiveness research. Qual Life Res. 2013;22(8):1889–905.

26. Ebesutani C, Reise SP, Chorpita BF, Ale C, Regan J, Young J, et al. The Revised Child Anxiety and Depression Scale-Short Version: Scale reduction via exploratory bifactor modeling of the broad anxiety factor. Psychol Assess. 2012;24(4):833–45.

27. Chorpita BF, Yim L, Moffitt C, Umemoto LA, Francis SE. Assessment of symptoms of DSM-IV anxiety and depression in children: A revised child anxiety and depression scale. Behav Res Ther. 2000;38(8):835–55.

28. Piqueras JA, Martín-Vivar M, Sandin B, San Luis C, Pineda D. The Revised Child Anxiety and Depression Scale: A systematic review and reliability generalization meta-analysis. J Affect Disord. 2017;218:153–69.

29. Foa EB, Coles M, Huppert JD, Pasupuleti R V., Franklin ME, March J. Development and Validation of a Child Version of the Obsessive Compulsive Inventory. Behav Ther. 2010;41(1):121–32.

30. Yule W. Anxiety, Depression and Post-Traumatic Stress in Childhood. In: Sclare I, editor. Child Psychology Portfolio. Windsor, England: NFER-Nelson; 1997.

31. Posner K, Brown GK, Stanley B, Brent DA, Yershova K V., Oquendo MA, et al. The Columbia-suicide severity rating scale: Initial validity and internal consistency findings from three multisite studies with adolescents and adults. Am J Psychiatry. 2011;168(12):1266–77.

32. Hoagwood KE, Jensen PS, Petti T, Burns BJ. Outcomes of mental health care for children and adolescents: I. A comprehensive conceptual model. J Am Acad Child Adolesc Psychiatry. 1996;35(8):1055–63.

33. Rapee RM, Bögels SM, Van Der Sluis CM, Craske MG, Ollendick T. Annual research review: Conceptualising functional impairment in children and adolescents. J Child Psychol Psychiatry Allied Discip. 2012;53(5):454–68.

34. Ravens-Sieberer U, Erhart M, Rajmil L, Herdman M, Auquier P, Bruil J, et al. Reliability, construct and criterion validity of the KIDSCREEN-10 score: A short measure for children and adolescents' well-being and health-related quality of life. Qual Life Res. 2010;19(10):1487–500.

35. Shaffer D, Gould MS, Brasic J, Fisher P, Aluwahlia S, Bird H. A Children's Global Assessment Scale (CGAS). Arch Gen Psychiatry. 1983;40(11):1228–31.

36. Lyneham HJ, Sburlati ES, Abbott MJ, Rapee RM, Hudson JL, Tolin DF, et al. Psychometric properties of the Child Anxiety Life Interference Scale (CALIS). J Anxiety Disord. 2013;27(7):711–9.

37. UNESCO Institute of Statistics. International Standard Classification of Education: ISCED 2011 [Internet].

Montreal, PQ, Canada; 2012. Available from: http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-isced-2011-en.pdf [18/06/2020].

38. Southam-Gerow MA, Kendall PC, Weersing VR. Examining outcome variability: Correlates of treatment response in a child and adolescent anxiety clinic. J Clin Child Psychol. 2001;30(3):422–36.

39. Berman SL, Weems CF, Silverman WK, Kurtines WM. Predictors of outcome in exposure-based cognitive and behavioral treatments for phobic and anxiety disorders in children. Behav Ther. 2000;31(4):713–31.

40. Liber JM, Van Widenfelt BM, Van Der Leeden AJM, Goedhart AW, Utens EMWJ, Treffers PDA. The relation of severity and comorbidity to treatment outcome with cognitive behavioral therapy for childhood anxiety disorders. J Abnorm Child Psychol. 2010;38(5):683–94.

41. Emslie GJ, Kennard BD, Mayes TL. Predictors of treatment response in adolescent depression. Psychiatr Ann. 2011;41(4):212–9.

42. Lundkvist-Houndoumadi I, Hougaard E, Thastum M. Pre-treatment child and family characteristics as predictors of outcome in cognitive behavioural therapy for youth anxiety disorders. Nord J Psychiatry. 2014;68(8):524–35.

43. Brent DA, Kolko DJ, Birmaher B, Baugher M, Bridge J, Roth C, et al. Predictors of treatment efficacy in a clinical trial of three psychosocial treatments for adolescent depression. J Am Acad Child Adolesc Psychiatry. 1998;37(9):906–14.

44. Rapee RM, Lyneham HJ, Hudson JL, Kangas M, Wuthrich VM, Schniering CA. Effect of comorbidity on treatment of anxious children and adolescents: Results from a large, combined sample. J Am Acad Child Adolesc Psychiatry. 2013;52(1):47–56.

45. Storch EA, Merlo LJ, Larson MJ, Geffken GR, Lehmkuhl HD, Jacob ML, et al. Impact of comorbidity on cognitive-behavioral therapy response in pediatric obsessive-compulsive disorder. J Am Acad Child Adolesc Psychiatry. 2008;47(5):583–92.

46. Skriner LC, Chu BC, Kaplan M, Bodden DHM, Bögels SM, Kendall PC, et al. Trajectories and predictors of response in youth anxiety CBT: Integrative data analysis. J Consult Clin Psychol. 2019;87(2):198–211.

47. Hudson JL, Keers R, Roberts S, Coleman JRI, Breen G, Arendt K, et al. Clinical predictors of response to cognitive-behavioral therapy in pediatric anxiety disorders: The genes for treatment (GxT) study. J Am Acad Child Adolesc Psychiatry. 2015;54(6):454–63.

48. Jones M, Hopkins K, Kyrke-Smith R, Davies R, Vostanis P, Wolpert M. Current view tool: Completion guide. London, United Kingdom: CAMHS Press; 2013.

49. World Health Organization. International Classification of Diseases for Mortality and Morbidity Statistics (11th Revision) [Internet]. 2018 [cited 2019 Oct 15]. Available from: https://icd.who.int/browse11/l-m/en [18/06/2020].

50. Bodden DHM, Bögels SM, Nauta MH, De Haan E, Ringrose J, Appelboom C, et al. Child versus family cognitive-behavioral therapy in clinically anxious youth: An efficacy and partial effectiveness study. J Am Acad Child Adolesc Psychiatry. 2008;47(12):1384–94.

51. Aggelosopoulou A, Roussi-Vergou C, Zafeiropoulou M. "Anxioussaure": A CBT psychoeducational programme for children's anxiety. Programme development and pilot study. Psychol. J. Hell. Psychol. Soc. 2012; 19(3):245–262

52. Legerstee JS, Huizink AC, van Gastel W, Liber JM, Treffers PDA, Verhulst FC, et al. Maternal anxiety predicts favourable treatment outcomes in anxiety-disordered adolescents. Acta Psychiatr Scand. 2008;117(4):289–98.

53. Nanni V, Uher R, Danese A. Childhood maltreatment predicts unfavorable course of illness and treatment outcome in depression: A meta-analysis. Am J Psychiatry. 2012;169(2):141–51.

54. Batty MJ, Moldavsky M, Foroushani PS, Pass S, Marriott M, Sayal K, et al. Implementing routine outcome measures in child and adolescent mental health services: From present to future practice. Child Adolesc Ment Health. 2013;18(2):82–7.

55. Chevance A, Tran VT, Ravaud P. Improving the generalizability and credibility of core outcome sets (COSs) by a large and international participation of diverse stakeholders. J Clin Epidemiol. 2020 Jan;1–8.

56. Davis K, Gorst SL, Harman N, Smith V, Gargon E, Altman DG, et al. Choosing important health outcomes for comparative effectiveness research: An updated systematic review and involvement of low and middle income countries. PLoS One. 2018;13(2):1–14.

57. Kieling C, Rohde LA. Child and Adolescent Mental Health Research Across the Globe. J Am Acad Child Adolesc Psychiatry. 2012;51(9):945–7.

58. Trauer T. Outcome Measurement in Mental Health: Theory and Practice. Cambridge, England: Cambridge University Press; 2010.

59. Delespaul PAEG. Routine outcome measurement in the Netherlands-A focus on benchmarking. Int Rev

Psychiatry. 2015;27(4):320–8.

60. Department of Health and Ageing. Mental Health National Outcomes and Casemix Collection: Overview of Clinician-Rated and Consumer Self-Report Measures V1.50 [Internet]. Canberra; 2003. Available from: https://www.amhocn.org/sites/default/files/publication_files/mhnocc_overview_doha.pdf [18/06/2020].

61. National Quality Forum. Measure Applications Partnership, 2019 [Internet]. 2019. Available from: https://www.qualityforum.org/Setting_Priorities/Partnership/Measure_Applications_Partnership.aspx [22/06/2020].

62. Monga S, Monsour A, Stallwood E, Desai R, Cleverley K, Courtney D, et al. Core Outcome Set Development for Adolescent Major Depressive Disorder Clinical Trials: A Registered Report. J Am Acad Child Adolesc Psychiatry. Forthcoming 2020.

63. UNICEF. Measurement of Mental Health Among Adolescents at the Population Level (MMAP): An Overview [Internet]. 2019. Available from: https://data.unicef.org/wp-content/uploads/2018/02/Mental-health-measurement_MMAP_overview_UNICEF.pdf [06/12/2019].

64. Myers K, Winters NC. Ten-Year Review of Rating Scales. II: Scales for Internalizing Disorders. J Am Acad Child Adolesc Psychiatry. 2002;41(6):634–59.

65. Winters NC, Collett BR, Myers KM. Ten-year review of rating scales, VII: Scales assessing functional impairment. J Am Acad Child Adolesc Psychiatry. 2005;44(4):309–38.

66. Becker-Haimes EM, Tabachnick AR, Last BS, Stewart RE, Hasan-Granier A, Beidas RS. Evidence Base Update for Brief, Free, and Accessible Youth Mental Health Measures. J Clin Child Adolesc Psychol. 2020;49(1):1–17.

67. Beidas RS, Stewart RE, Walsh L, Lucas S, Downey MM, Jackson K, et al. Free, brief, and validated: Standardized instruments for low-resource mental health settings. Cogn Behav Pract. 2015;22(1):5–19.

68. Mash EJ, Hunsley J. Evidence-based assessment of child and adolescent disorders: Issues and challenges. J Clin Child Adolesc Psychol. 2005;34(3):362–79.

69. Reeve BB, Thissen D, DeWalt DA, Huang I-C, Liu Y, Magnus B, et al. Linkage between the PROMIS® pediatric and adult emotional distress measures. Qual Life Res. 2016;25(4):823–33.

70. Sales CMD. Seeing the person in the patient: Making the case for individualized PROMs in mental health care. Curr Psychiatry Rev. 2017;13(3):184–7.

71. Klaufus L, Verlinden E, Van Der Wal M, Kösters M, Cuijpers P, Chinapaw M. Psychometric evaluation of two short versions of the Revised Child Anxiety and Depression Scale. BMC Psychiatry. 2020;20(1):1–12.

72. L. Milfont T, Fischer R. Testing measurement invariance across groups: applications in cross-cultural research. Int J Psychol Res. 2010;3(1):111–30.

73. Bagheri Z, Akbari M, Jafari P, Huedo-medina TB. Assessing the measurement invariance of the RCADS-25 questionnaire across gender and child–parent dyads in the presence of multilevel data. Qual Life Res. 2019;28(3):829–40.

74. Rodríguez-Jiménez T, Piqueras JA, Lázaro L, Moreno E, Ortiz AG, Godoy A. Metric invariance, reliability, and validity of the Child Version of the Obsessive Compulsive Inventory (OCI-CV) in community and clinical samples. J Obsessive Compuls Relat Disord. 2016;9:1–8.

75. Rosa-Alcázar AI, Ruiz-García B, Iniesta-Sepúlveda M, López-Pina JA, Rosa-Alcázar Á, Parada-Navas JL. Obsessive Compulsive Inventory-Child Version (OCI-CV) in a Spanish community sample of children and adolescents. Psicothema. 2014;26(2):174–9.

76. Martínez-González AE, Rodríguez-Jiménez T, Piqueras JA, Vera-Villarroel P, Godoy A. Psychometric properties of the obsessive-compulsive Inventory-Child Version (OCI-CV) in Chilean children and adolescents. PLoS One. 2015;10(8):1–12.

77. Jones AM, De Nadai AS, Arnold EB, McGuire JF, Lewin AB, Murphy TK, et al. Psychometric properties of the obsessive compulsive inventory: Child version in children and adolescents with obsessive-compulsive disorder. Child Psychiatry Hum Dev. 2013;44(1):137–51.

78. Opakunle T, Aloba O, Akinsulore A. Obsessive compulsive inventory – child version (OCI-CV): Confirmatory factor analysis, reliability, validity and correlates among nigerian adolescents. Malawi Med J. 2018;30(4):262–9.

79. Verlinden E, van Meijel EPM, Opmeer BC, Beer R, de Roos C, Bicanic IAE, et al. Characteristics of the children's revised impact of event scale in a clinically referred dutch sample. J Trauma Stress. 2014;27(3):338–44.

80. Magalhães S de S, Miranda DK, de Paula JJ, de Miranda DM, Romano-Silva MA, Malloy-Diniz LF. Psychometric properties of a Brazilian portuguese version of the children's revised impact of event scale (CRIES-8). Rev Psiquiatr Clin. 2018;45(2):27–32.

81. Salari R, Malekian C, Linck L, Kristiansson R, Sarkadi A. Screening for PTSD symptoms in unaccompanied

refugee minors: A test of the CRIES-8 questionnaire in routine care. Scand J Public Health. 2017;45(6):605–11.

82. Gigengack MR, Hein IM, Lindeboom R, Lindauer RJL. Increasing Resource Parents' Sensitivity towards Child Posttraumatic Stress Symptoms: a Descriptive Study on a Trauma-Informed Resource Parent Training. J Child Adolesc Trauma. 2019;12(1):23–9.

83. Kilincaslan A, Gunes A, Eskin M, Madan A. Linguistic adaptation and psychometric properties of the Columbia-Suicide Severity Rating Scale among a heterogeneous sample of adolescents in Turkey. Int J Psychiatry Med. 2019;54(2):115–32.

84. Nik-Azin A, Shairi MR, Naeinian MR, Sadeghpour A. The health-related quality of life index kidscreen-10: Confirmatory factor analysis, convergent validity and reliability in a sample of Iranian students. Child Indic Res. 2014;7(2):407–20.

85. Stevanovic D, Tadic I, Novakovic T, Kisic-Tepavcevic D, Ravens-Sieberer U. Evaluating the Serbian version of the KIDSCREEN quality-of-life questionnaires: Reliability, validity, and agreement between children's and parents' ratings. Qual Life Res. 2013;22(7):1729–37.

86. Weissman MM, Warner V, Fendrich M. Applying Impairment Criteria to Children's Psychiatric Diagnosis. J Am Acad Child Adolesc Psychiatry. 1990;29(5):789–95.

87. Bird HR. Further Measures of the Psychometric Properties of the Children's Global Assessment Scale. Arch Gen Psychiatry. 1987;44(9):821.

88. Dyrborg J, Warborg Larsen F, Nielsen S, Byman J, Buhl Nielsen B, Gautrè-Delay F. The Children's Global Assessment Scale (CGAS) and Global Assessment of Psychosocial Disability (GAPD) in clinical practice - Substance and reliability as judged by intraclass correlations. Eur Child Adolesc Psychiatry. 2000;9(3):195–201.

89. Sourander A, Helenius H, Piha J. Child psychiatric short-term inpatient treatment: CGAS as follow-up measure. Child Psychiatry Hum Dev. 1996;27(2):93–104.

90. Orgilés M, Fernández-Martínez I, Morales A, Melero S, Espada JP. Spanish Validation of the Child Anxiety Life Interference Scale (CALIS-C): Psychometric Properties, Factorial Structure and Factorial Invariance Across Gender. Child Psychiatry Hum Dev. 2019;50(5):756–63.

91. Marques T, Pereira AI, Pedro M, Russo V, Goes AR, Barros L. Preliminary analysis of the Portuguese version of the child anxiety life interference scale. Psicol Reflex e Crit. 2015;28(3):454–62.

Accepted for publication by the Lancet Psychiatry

Appendix

Measuring Response to Clinical Care in Children and Young People with Anxiety, Depression, OCD or PTSD: An International Standard Set of Outcome Measures

Karolin Krause, Sophie Chung, Abiodun O. Adewuya, Anne Marie Albano, Rochelle Babins-Wagner, Laura Birkinshaw, Peter Brann, Cathy Creswell, Kathleen Delaney, Bruno Falissard, Christopher B. Forrest, Jennifer L. Hudson, Shin-ichi Ishikawa, Meghna Khatwani, Christian Kieling, Judi Krause, Kanika Malik, Vania Martínez, Faraz Mughal, Thomas H. Ollendick, Say How Ong, George C. Patton, Ulrike Ravens-Sieberer, Peter Szatmari, Evie Thomas, Lucas Walters, Bryan Young, Yue Zhao, Miranda Wolpert

Corresponding Author: Karolin Krause, Research Department of Clinical, Educational and Health Psychology, University College London, Gower Street, London WC1E 6BT, United Kingdom; email: karolin.krause.16@ucl.ac.uk.

Table of Contents

Table of Tables

Table of Figures

Notes on the Scope of the Standard Set

As per Working Group consensus, this Standard Set was designed for use with children and young people (CYP) aged 6-24 years, with anxiety, depression, obsessive-compulsive disorder (OCD), or post-traumatic stress disorder (PTSD) as defined by the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5; 1), or the International Classification of Diseases for Mortality and Morbidity Statistics, 11th Revision (ICD-11; 2). All included conditions are internalising disorders, typically characterised by high levels of negative affect (3). Depression and anxiety have high levels of co-occurrence (4,5); and PTSD and OCD were classified as anxiety disorders until the publication of the DSM-5 (3). Within the broader diagnostic categories of depression and anxiety, all relevant sub-disorders are covered (e.g., separation anxiety, panic disorder, or social phobia), with the exception of secondary or substance-induced depressive or anxiety disorders, and selective mutism. Bipolar disorder is excluded. OCD and PTSD are included as primary diagnoses, but other OCD- and stress-related disorders are not explicitly covered (e.g., body focused repetitive behaviour disorder, complicated bereavement). This means that the Working Group did not seek to identify outcomes or measurement instruments that were specifically applicable to these sub-disorders. However, services and practitioners are encouraged to use this Standard Set for CYP with other presentations as and where relevant.

The lower age threshold of 6 years was chosen to reflect differences in the presentation, treatment, and monitoring of the target disorders in young children, as well as cross-cultural variation in the minimum age for diagnoses. Similarly, definitions of adolescence vary cross-culturally, as does the typical age of transition into adult care. An existing ICHOM Standard Set for adult anxiety and depression covers young people from the age of 14 years (6). The combination of the Sets for adults and CYP provides for transition at any point between the ages of 14 and 24, allowing for local variation in transition and judgements about which Set is most suitable for different ages. To avoid variation *within* services, the Working Group recommends that CYP's mental health services generally administer the Set for CYP, while adult services should administer the ICHOM Standard Set for Depression and Anxiety in Adults. However, responsibility for selecting the most suitable set will ultimately lie with clinicians.

Detailed Methodology

The Working Group

The Working Group included six experts by experience (i.e., young people with experience of service use for anxiety, depression, OCD, or PTSD; or their parents), and 21 experts by profession, including biostatisticians, clinical psychologists, epidemiologists, nurses, primary care physicians, psychiatrists, psychometricians, and social workers. Working Group members came from 13 countries across Africa, the Americas, Asia, Europe, and Oceania (table S1). In addition, a central project team (KK, SC, MW) coordinated and facilitated the consensus-building process, and completed supporting research tasks, but did not vote with the Working Group on the recommendations to be made. This included KK as research fellow, SC as project manager, and MW as Working Group chair. The non-voting project team were all based in the United Kingdom.

Table S1. Composition of the Working Group (27 voting members)

Region / Subregion	***n (%)***
Americas	10 (37%)
North America	8 (30%)
Latin America & the Caribbean	2 (7%)
Africa	1 (4%)
Northern and Sub-Saharan Africa	1 (4%)
Asia	6 (22%)
Eastern and Southern Asia	6 (22%)
Europe	5 (19%)
Northern Europe	3 (11%)
Western Europe	2 (7%)
Oceania	5 (19%)
Total	27 (100%)

Working Group members were identified through several avenues. Firstly, a scoping review was conducted by SC during the project initiation phase, to identify relevant service user organisations, measurement initiatives, professional bodies, and publications actively addressing questions relating to outcome measurement for anxiety and depression in CYP. In addition, open recruitment calls inviting interested individuals to participate in the Working Group, Open Review process, or Standard Set implementation, was disseminated through ICHOM's networks, including the funders of this initiative, as well as social media channels. Relevant organisations identified as part of the rapid review were contacted and the information shared. Individuals were identified through both routes. In the first instance, ICHOM identified a potential Working Group chair (MW), who was subsequently contacted and engaged. Next, a matrix of candidates was composed to facilitate the representation of diverse geographies, disciplines, types of expertise (i.e., professional versus lived experience), as well as a balance of specialist interests (i.e. in depression or anxiety, or in specific developmental groups). A shortlist was created that would represent different matrix cells, and shortlisted professionals were invited to participate by ICHOM. There was a snow-ball sampling element to this process, whereby individuals or organisations could recommend additional candidates for consideration by ICHOM, or agree to disseminate the recruitment call (7). Young people and parents with lived experience of service use were invited based on their pro-active solicitation of ICHOM in response to the recruitment calls disseminated via the above-mentioned channels. All participants were briefed on the project before agreeing to be part of the Working Group and all Working Group members participated on a voluntary basis. Of the six lived experience experts who joined the Working Group, three participated throughout the entirety of the process, and three participated throughout the Delphi phases and endorsed the final set. There was no drop-out amongst the 21 professional experts who joined the Working Group process from the outset.

The Consensus-Building Process

Consensus was built through structured teleconferences, a Delphi-type exercise, and iterative rounds of anonymous feedback surveys and voting, which were informed by systematic and narrative literature reviews and other research inputs completed by the central project team. An open online consultation (henceforth called *Open Review*) gathered external feedback on the draft recommendations from researchers, practitioners, and service user representatives beyond the Working Group, toward the end of the consensus-building process.

Structured teleconferences. The Working Group convened for eight thematic structured teleconferences over a period of 14 months (October 2018 – December 2019). Professional Working Group members were required to participate in a minimum of 50% of the teleconferences and associated surveys to be considered full Working Group members. Lived-experience experts were expected to participate in the Delphi exercise, and strongly encouraged to also participate in all other stages of the process. On average, each teleconference was attended by 17 of the 27 Working Group members (ranging from 14 to 21 attendees across the eight calls), who typically split across two separate calls to accommodate different time zones. Catch-up calls were offered to all members unable to make the full call. During each teleconference, the central project team (KK, SC, MW) presented research findings (e.g., from the systematic review or the appraisal of measurement instruments, see below), as well as anonymous voting results and comments from the Delphi exercise and feedback surveys (see below). Key decision points were suggested by the project team and discussed within the group. As part of this, the Working Group chair (MW) would invite all participants, one participant at a time, to share their thoughts on the decision to be made in a 'round robin' fashion, with a different order selected each time. Participants also had the opportunity to raise additional issues or questions or to come back to points raised at the end of each round. No formal decisions were made during teleconferences and full minutes were shared with all Working Group members.

Iterative internal feedback surveys and online voting. All decisions about the Standard Set's scope, included outcome domains, and recommendations in relation to measurement instruments, case-mix factors, and timepoints were based on iterative voting via anonymous online surveys. After each teleconference, a survey was circulated for members to vote on the key decision points identified during the call. Summaries of the research findings and discussion points from the corresponding conference call were provided for general reference, and to inform those who had been unable to attend the call. At least 80% of Working Group members had to vote in any given Working Group survey for the results to be considered valid; and every decision had to be endorsed by at least 70% of survey participants for consensus to be considered as reached. The central project team (KK, SC, MW) had a facilitating role and did not participate in the voting. Where required, initial voting results and free text comments were shared within the Working Group during the subsequent call, to facilitate movement towards consensus in a subsequent round of voting. In some cases, the least favoured option from prior rounds was dropped, with the process made clear prior to

voting. Areas of contention were openly discussed, and anonymous voting repeated until consensus was achieved. This structured process was used to reach consensus on the scope of the Standard Set, suitable outcome measures, case-mix variables, and time points for measurement. The Working Group also voted to confirm their support on procedural aspects, such as the criteria for instrument selection.

Delphi-type exercise to select outcome domains. In order to determine the outcome domains for inclusion, a modified 3-round Delphi process was run following the second teleconference. As in all Working Group surveys, 80% participation was required in each round. In the first round, each voting Working Group member ranked each potential outcome domain on a scale from 1 to 9 based on a number of criteria, to indicate whether they thought the domain should be included in the Set. This was in line with common practice in Delphi surveys, were 9-point rating scales are frequently used (8). Domains ranked between 7 and 9 by 80% of the Working Group were included after round one. Those domains that were ranked between 1 and 3 by 80% of the Working Group were dropped. For all outcome domains that fell in between, written comments provided by Working Group members in the Delphi survey were shared. A second round of voting with the same process was then completed. After the third teleconference, all remaining ambiguous outcome domains were discussed during a Working Group call and then subject to a *Yes/No* inclusion vote, with 70% consensus required for inclusion. As described above, decisions about the Standard Set's scope, measurement instruments, case-mix factors, and time points were reached using simplified voting techniques and rating scales, including binary response options, and choices between multiple alternative options.

Research Inputs

Systematic Review. A systematic literature review was conducted at the start of the process to identify an exhaustive list of possible outcomes, measurement instruments, and case mix factors.

Search Strategy. A search syntax was devised to identify relevant studies in three publication databases: Medline, APA PsychINFO, and the Cumulative Index to Nursing and Allied Health Literature (CINAHL). Rather than searching for pre-defined outcome concepts, the project team searched for a wide range of treatment efficacy and effectiveness studies, to identify the outcomes measured and reported in the empirical literature. The search terms specified the study population (i.e., children, adolescents, young people), the target disorders (i.e., anxiety, depression, OCD, PTSD), and a wide range of methodologies (e.g., clinical trials, observational studies, case series). The search syntax further included terms to narrow down the search to treatment outcome studies (e.g., treatment effects, intervention efficacy, outcome assessment), and a number of exclusion terms (e.g., feasibility study). The search strategy used to search APA PsychINFO is provided in table S2.

Inclusion and exclusion criteria. To be considered for inclusion, studies had to (a) be published between January 2013 and November 2018 in a peer-reviewed journal, (b) focus on treatment for anxiety, depression, OCD, or PTSD in CYP aged 6-24 years, with a mean participant age within the eligible age range, (c) enrol participants who had received a clinical diagnosis or referral, or were seeking help for one of the target disorders (i.e., prevention studies were excluded), (d) assess treatment outcome quantitatively for at least one of the target disorders, and (e) be published in English, French, German, or Spanish. Eligible study designs included randomised control trials (RCTs) and follow-ups, open-label trials and naturalistic studies, case-control studies and case series. In the case of trials, trial protocols were reviewed in addition to original research articles to identify any outcomes and measures that had been included in the original study design but were not subsequently reported. No restrictions were imposed in relation to study settings or intervention modalities.

Study Selection. The systematic search identified 1,562 unique articles. Titles and abstracts were split into three parts and screened by two members of the central project team (KK and SC) and one project assistant (CI). Of the 1,562 articles screened, 418 were retained for full-text screening (figure s2). These were split evenly between the three reviewers, and 10% of full-texts were screened by all three reviewers. A *Fleiss Kappa* of 0.70 indicated good interrater agreement between the three reviewers (8). Discrepancies were discussed and resolved. A total of 257 articles were retained for data extraction.

Data Extraction. Data was systematically extracted by two reviewers (KK and SC) using a data extraction matrix that had been piloted in a previous systematic review of outcomes measured in treatment outcome studies for adolescent depression (9). At the start of the data extraction process, the two reviewers extracted data in duplicate from ten studies. Results were compared, discrepancies discussed, and the extraction matrix refined further to enhance data quality and

consistency. The remainder of studies were divided between SC and KK for data exaction. Information extracted from each study included: citation details, trial registration details (where applicable), country of implementation, research design, target disorder and sub disorder(s), information about the participant sample (i.e., gender ratio, age range, mean age, ethnicity, sample size), the eligibility criteria applied for participant recruitment (i.e., clinical diagnosis, help-seeking, referral), exclusion criteria, intervention setting, intervention type, moderators and predictors of treatment assessed in the study, the outcome concepts assessed, the outcome measure used, and the informants consulted on each outcome. All outcomes and measures identified were systematically catalogued. As this review aimed to identify outcome concepts and measurement instruments, rather than to synthesise the evidence base on treatment efficacy or effectiveness, the 257 studies included were not subject to a data quality assessment.

Data synthesis. The aim of data synthesis was to collate a maximum number of meaningfully distinct outcome concepts for Working Group members to vote on in the Delphi exercise. One reviewer (SC) reviewed all primary and secondary outcome concepts that had been extracted 'verbatim' from the included studies, and collapsed duplicates (e.g., 'symptoms of anxiety' and 'anxiety symptoms'), as well as significantly overlapping concepts (i.e. 'psychosocial functioning' and 'social functioning'). These decisions were made with reference to existing conceptual outcome frameworks (10,11) and reviewed by a second rater (KK). For certain concepts, the Working Group deliberated on the final terminology to describe the outcome concept. Where possible, consistency was maintained with prior ICHOM Standard Sets. In addition, all outcome measurement instruments reported in the reviewed studies were extracted and added to a database that mapped each instrument to the primary concept(s) that they were designed to measure according to key validation papers. Once the Working Group had agreed on a broad range of outcomes to consider, all instruments in the database were mapped onto these outcome domains by two members of the central project team (KK and SC), who reviewed and discussed each other's domain mapping iteratively, to enhance reliability and consistency.

Table S2. Search Strategy for Systematic Review (PsychINFO)

Search Terms
1. "0300".md.
2. randomized.ab.
3. placebo.ab.
4. (clinical trial or randomi#ed control# trial).id.
5. randomly.ab.
6. trial.ti.
7. 1 or 2 or 3 or 4 or 5 or 6
8. exp animals/ not humans.sh.
9. (primate* or rat#1 or mouse or mice or non-human).ti.
10. 8 or 9
11. 7 not 10
12. ("0430" or "0452" or "0600" or "1600").md.
13. (case$ and control$).tw.
14. (case$ and series).tw.
15. observational study.ab,ti,mh.
16. 12 or 13 or 14 or 15
17. (case study or case report or single subject).ti.
18. 16 not 17
19. (("semi-structured" or semistructured or unstructured or informal or "in-depth" or indepth or "face-to-face" or structured or guide) adj3 (interview* or discussion* or questionnaire*)).ti,ab.
20. (focus group* or qualitative or ethnograph* or fieldwork or "field work" or "key informant").ti,ab.
21. interviews as topic/ or focus groups/ or narration/ or qualitative research/
22. 19 or 20 or 21
23. (child*3 or kids or childhood or schoolchild* or pupil* or school-age* or school age*).ti.
24. (adoles* or teen* or boy* or girl* or minors or puberty or youth or juvenile*).ti.
25. ((young adj person) or (young adj people)).ti.
26. (paediatric* or pediatric*).ti.
27. school*.ti.
28. (adolescent or child or minors).mh.
29. 23 or 24 or 25 or 26 or 27 or 28
30. adult.mh.
31. 29 not 30
32. (depress* or mdd or antidepress* or (low adj3 mood) or (mood adj3 disorder*) or (affective adj3 disorder*) or (internali#ing adj3 disorder*) or suicid*).ti.
33. (cyclothymic disorder* or dysthymic disorder* or seasonal affective disorder* or single episode depress* or recurrent depress* or premenstrual dysphoric disorder or disruptive mood or (mood adj dysregulat*)).ti.

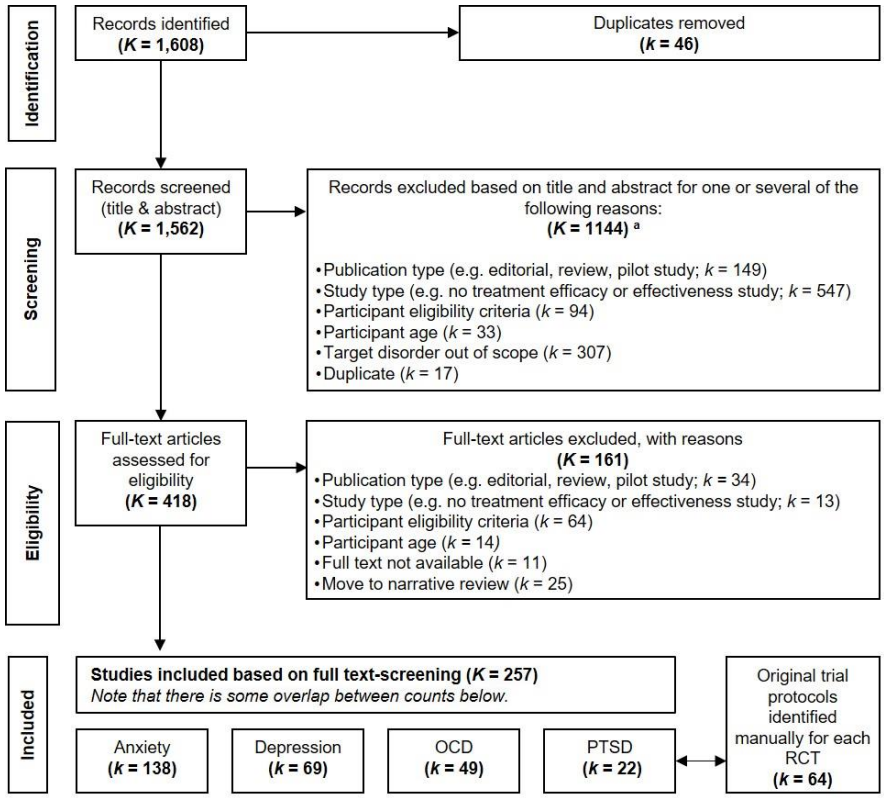
Search Terms
34. (anxiety or panic or phobi* or agoraphobia).ti.
35. (obsessive#compulsive or obsession* or obsessive* or compulsion* or compulsive*).ti.
36. (post#traumatic or traumatic stress or ptsd).ti.
37. (depression or anxiety or ptsd or post#traumatic stress or ocd or obsessive#compulsive disorder).id.
38. (depression or affective symptoms or anxiety disorders or mood disorders or obsessive-compulsive disorder or stress disorders, traumatic).mh.
39. 32 or 33 or 34 or 35 or 36 or 37 or 38
40. ((intervention adj6 effect*) or (intervention adj6 efficacy) or (intervention adj6 outcome*) or (intervention adj6 success) or (intervention adj6 impact) or (intervention adj6 result*) or (intervention adj6 evaluat*) or (intervention adj6 measu*)).ab,ti.
41. ((intervention adj10 experience*) or (intervention adj10 perception*) or (intervention adj10 descri*) or (intervention adj10 feedback*)).ab,ti.
42. ((treat* adj6 effect*) or (treat* adj6 efficacy) or (treat* adj6 outcome*) or (treat* adj6 success) or (treat* adj6 impact) or (treat* adj6 result*) or (treat* adj6 response) or (treat* adj6 evaluat*) or (treat* adj6 measu*)).ab,ti.
43. ((treat* adj10 experience*) or (treat* adj10 perception*) or (treat* adj10 descri*) or (treat* adj10 feedback*)).ab,ti.
44. ((program*3 adj6 effect*) or (program*3 adj6 efficacy) or (program*3 adj6 outcome*) or (program*3 adj6 success) or (program*3 adj6 impact) or (program*3 adj6 result*) or (program*3 adj6 evaluat*)).ab,ti.
45. ((program*3 adj10 experience*) or (program*3 adj10 perception*) or (program*3 adj10 descri*) or (program*3 adj10 feedback*)).ab,ti.
46. ((psychotherap* adj6 effect*) or (psychotherap* adj6 efficacy) or (psychotherap* adj6 outcome*) or (psychotherap* adj6 success) or (psychotherap* adj6 impact) or (psychotherap* adj6 result*) or (psychotherap* adj6 evaluat*) or (psychotherap* adj6 measu*)).ab,ti.
47. ((therap* adj6 effect*) or (therap* adj6 efficacy) or (therap* adj6 outcome*) or (therap* adj6 success) or (therap* adj6 impact) or (therap* adj6 result*) or (therap* adj6 evaluat*)).ab,ti.
48. ((psychotherap* adj10 experience*) or (psychotherap* adj10 perception*) or (psychotherap* adj10 descri*) or (psychotherap* adj10 feedback*)).ab,ti.
49. ((therap* adj10 experience*) or (therap* adj10 perception*) or (therap* adj10 descri*) or (therap* adj10 feedback*)).ab,ti.
50. ((evaluat* adj6 effect*) or (evaluat* adj6 efficac*) or (evaluat* adj6 success) or (evaluat* adj6 response) or (evaluat* adj6 outcome*) or (evaluat* adj6 result*) or (evaluat* adj6 pretreat*) or (evaluat* adj6 posttreat*)).ab,ti.
51. ((assess* adj6 effect*) or (assess* adj6 efficac*) or (assess* adj6 success) or (assess* adj6 response) or (assess* adj6 outcome*) or (assess* adj6 result*) or (assess* adj6 pretreat*) or (assess* adj6 posttreat*)).ab,ti.
52. ((measur* adj6 effect*) or (measur* adj6 efficac*) or (measur* adj6 success) or (measur* adj6 response) or (measure* adj6 outcome*) or (measur* adj6 result*) or (measur* adj6 pretreat*) or (measur* adj6 posttreat*)).ab,ti.
53. ((trial adj6 outcome*) or (trial adj6 effect*)).ab,ti.
54. (outcome assessment or outcome measurement).id.
55. (treatment outcome or patient outcome assessment or patient reported outcome measures or clinical trial or evaluation studies).mh.
56. 40 or 41 or 42 or 43 or 44 or 45 or 46 or 47 or 48 or 49 or 50 or 51 or 52 or 53 or 54 or 55
57. ("2018" or "2017" or "2016" or "2015" or "2014" or "2013").yr.
58. ("0100" or "0110" or "0120").pt.
59. prevent*.ti.
60. ((pilot adj3 study) or (pilot adj3 trial) or (pilot adj3 evaluation)).ti.
61. ((feasibility adj3 study) or (feasibility adj3 trial) or (feasibility adj3 evaluation)).ti.
62. ((protocol adj6 study) or (protocol adj6 trial)).ti.
63. (case study or case report or single subject study).ti.
64. 59 or 60 or 61 or 62 or 63
65. (injury or injuries or cancer or asthma or eczema or diabetes).ti.
66. ((anxiety adj5 preoperative) or (stress adj5 preoperative) or (anxiety adj5 perioperative) or (stress adj5 perioperative) or surgery or surgical).ti.
67. 65 or 66
68. 64 or 67
69. 11 or 18 or 22
70. 69 and 56
71. 70 and 31 and 39 and 57 and 58
72. 71 not 68

Note. Equivalent syntaxes were used for searches in Medline and CINAHL, with small syntax adjustments to meet the requirements of the relevant search engines.

Identification of outcome domains. An extensive list of possible outcome domains was compiled by the project team from the 257 studies identified through the systematic review. This was complemented with a narrative review of supplemental sources, including clinical guidelines, instrument banks, qualitative studies, and cohort studies. These were identified through manual searches of reference lists, and recommendations from Working Group members. In addition, break-out groups were conducted with the group's service user representatives to identify any additional outcomes, and to highlight those considered most important. After consolidating findings from the systematic and narrative reviews, and the consultation of Working Group members, 70 outcome domains were identified. Through the Delphi exercise (see above), this list was then gradually reduced; and consensus was formed on the outcome domains to include in the Standard Set. The aim was to seek a balance between tracking response to treatment comprehensively, and devising a feasible set of recommendations that services could implement reliably. To this end,

the Working Group were advised to identify a *core* Standard Set of up to 15 outcomes considered most relevant to children, young people, and their families, and which would be recommended as *a minimum*.

Figure S1. Modified PRISMA Diagram



[a] Note that *k* does not add up to *K* due to double coding for different exclusion criteria at the title and abstract screening stage.

Identification and appraisal of outcome measures. After reaching consensus on the outcome domains to include, the Working Group proceeded to the identification and appraisal of measurement instruments that would be most suitable for tracking change in the selected outcome domains over time. The systematic and narrative literature reviews identified 507 unique outcome measures, which included both clinician-rated and patient- reported instruments. After removing ad-hoc measures and tools not suitable for routine use (e.g., lab-based tests), 355 measurement instruments were retained and mapped to the outcome concept they were designed to measure, according to key development or validation papers (figure S2, below).

A Three-Stage Appraisal Process. Of these 355 measurement instruments, 107 broadly mapped onto one of the outcome domains selected for inclusion by the Working Group, and were extracted and subjected to a systematic, three-stage appraisal (see tables S3-S7). At stage one, the central project team (KK, SC, MW) determined whether each measure comprehensively covered the designated outcome domain. This process served to exclude measures that at closer assessment did not fall within any of the designated domains (e.g., Self-Efficacy Questionnaire for School Situations), or that only captured a specific aspect of the broader outcome concept (e.g., the Existential Anxiety Questionnaire). The majority of measures provided adequate domain coverage and were carried forward to stage two. At stage two, the instruments' feasibility and acceptability for clinical practice was assessed by establishing whether a measure (a) had been tested for reliability and validity in children or young people (as opposed to being developed for and validated in adults only); (b) was available at no licensing cost for clinical practice in paper format and as part of electronic systems; (c) was sufficiently brief (i.e., requiring less than 20 minutes for completion, or including fewer than 60 items); and (d) was available in more than one language (i.e., demonstrating a level of international transportability).

At stage three, studies meeting feasibility criteria were assessed against the minimum standards for patient-reported outcome measures suggested by the International Society for Quality of Life Research (ISOQOL; 12). Measures were evaluated in terms of whether or not they demonstrated acceptable test-retest or interrater reliability ($r \geq 0.7$) and internal consistency (Cronbach's $\alpha \geq 0.7$), and whether there was evidence of their sensitivity to change in terms of changes in scores measured overtime that were consistent with predefined hypotheses about the expected treatment outcome (12). Other aspects such as content and construct validity were also considered and discussed by the Working Group. Relevant studies reporting on psychometric properties were compiled through an automated search using a validated PubMed search filter (13), the consultation of an online database of outcome measures (https://eprovide.mapi-trust.org), and consideration of systematic reviews focusing on the psychometric properties of specific measures.

Results from the psychometric appraisal were first shared and discussed with a sub-set of interested Working Group members during domain-specific break-out groups (open to all Working Group members), which led to the shortlisting of the instruments that were judged to best satisfy the appraisal criteria. In the next teleconference, the shortlisted measures were presented along with the results from the psychometric appraisal, and a rationale for shortlisting them was provided. The psychometric appraisal results for measures that had not been shortlisted by the thematic break-out groups were made available to the Working Group participants as part of the appendix to the teleconference slides. Since few measures clearly exceeded all ISOQOL criteria (especially the relatively loosely defined criterion pertaining to sensitivity to change), the psychometric strengths and disadvantages of the shortlisted measures were discussed in detail within the Group. During these discussions, the Working Group considered additional aspects such as the length of the measure, its face validity, its readability, the age range covered by the measure, and whether matching youth and parent versions were available. The Working Group then voted in several iterations of online surveys to reach consensus on the measures to be included. The final selection was thus made by group deliberation and consensus, rather than through a criterion- or algorithm-based ranking process. As in all decisions relating to the Standard Set's recommendations, the central project team had a facilitating role but did not vote on the inclusion of outcome measures.

Working Group members had the opportunity to "lifeboat" any measure excluded by the process or by one of the thematic break-out groups at any of the three stages. This enabled individual Working Group members to review and challenge the criteria, to present a rationale for reconsidering the respective instrument, and to ensure further group discussion, voting, or a psychometric appraisal took place as needed to solidify the decision to be made.

The three-stage appraisal process considered both long and short versions of each measure. Following the feasibility assessment, evidence for the psychometric properties of short and long forms was reviewed. For the purpose of reducing the list of eligible measures to a shortlist, the thematic break-out groups considered evidence relating to short and long versions of each measure, with long forms often more widely validated. In the subsequent conference call with the full Working Group, the central project team then provided an overview of the psychometric appraisal results for all versions available for each measure shortlisted by the break-out groups. The Working Group would then make a final decision about the inclusion of short or long forms.

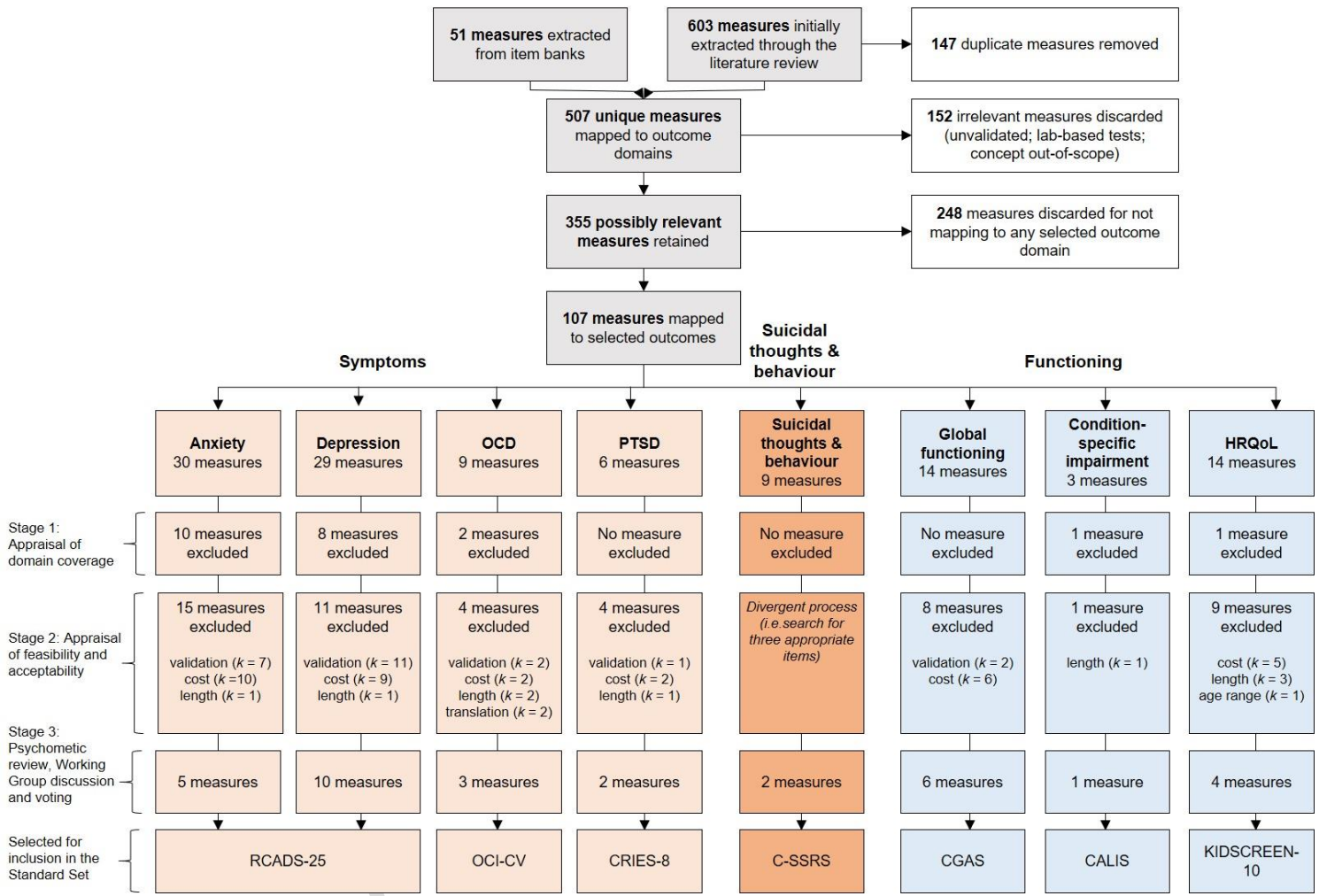
Process adaptation. The consensus building process aimed to yield a coherent set of recommendations for a feasible Standard Set. Consequently, discussions about each measure were, by necessity, not independent of one another. After reaching consensus on the most suitable symptom measures, the Working Group subsequently explored the most complementary options for measuring functioning and suicidal ideation and behaviour, with a view to minimising item redundancies and managing the overall length of the final instrument battery.

In this process, none of the self-reported measures of global functioning reached group consensus. This was due primarily to their length, and to the frequent conflation of symptom and impairment ratings, which led to overlap and redundancies with the selected symptom measures for anxiety and depression. A decision was made to explore disorder-specific impairment measures, and to consider complementing this with a shorter measure of Health-Related Quality of Life. Quality of life has been described as a more global and subjective construct than functional impairment, which "reflects the overall positivity with which individuals view their state and circumstances" across aspects of physical, psychological, cognitive, and social functioning (14). The Working Group considered that there was sufficient conceptual overlap to measure aspects of global functioning via a measure of HRQoL. An expedited review process took place to review measures originally mapped from the systematic review and supplemental sources (e.g., recommendations by the Working Group). Candidate measures were discussed at length in the teleconference calls, and through an additional break-out group.

After selecting the RCADS as a measure of anxiety and depression symptoms, the Working Group was not satisfied that this measure adequately covered the 'suicidal thoughts and behaviour' domain. However, in light of the length of available standalone measures of suicidality, the Group elected to cover this domain with a short set of items. All measures identified through the systematic review were reviewed and any short measures (up to three items), or individual items relating to suicidal thoughts and behaviour were extracted. In addition, the project ream reviewed supplemental sources (e.g., health surveys) and solicited suggestions from the Working Group. For any candidate measure or set of items, aspects relating to the coverage of suicidal thoughts *and* behaviour, cross-cultural applicability, and the reporter of this domain were discussed at length.

Figure S2 (below) provides an overview of the measure appraisal process. Tables S3 through S7 list the measurement tools identified within the domains of symptoms and functioning, and indicates the appraisal stage at which these measures were excluded (or included).

Figure S2. Flow Chart: Identification and Appraisal of Outcome Measures



[a] Shortlisted measures were reviewed against the minimum standards for patient-reported outcome measures recommended by the International Society for Quality of Life Research (ISOQOL; 12). Where long and short versions of a measure were available, all were considered for the ISOQOL assessment. Note that k does not always add up to 100%, since measures could be excluded for more than one reason. Note that seven measures were considered for both anxiety and depression, in addition to 23 anxiety measures and 24 depression measures.

Table S3. Overview of all Anxiety Measures Considered

Acronym	Measure Name	Comprehensive Domain Coverage	Feasibility Assessment	Subject to Psychometric Review	Included per Group Consensus
Measures Covering Anxiety & Depression					
CEMS	Children's Emotion Management Scales (15)	not passed	—	—	—
DASS-21	Depression Anxiety Stress Scale (16)	passed	not passed	—	—
HADS	Hospital Anxiety and Depression Scale (17)	passed	not passed	—	—
NASSQ	Negative Affect Self-Statement Questionnaire (18)	passed	passed	yes	no
PANAS	Positive and Negative Affect Scale for Children (19)	not passed	—	—	—
RCADS	Revised Child Anxiety and Depression Scales (20)	passed	passed	yes	yes
YSR	Youth Self-Report (21)	passed	not passed	—	—
Measures Covering Anxiety Only					
ADIS	Anxiety Disorders Diagnosis Interview Schedule DSM IV/V (22)	passed	not passed	—	—
BAI	Beck Anxiety Inventory (23)	passed	not passed	—	—
BFNE-II	Brief Fear of Negative Evaluation II (24)	not passed	—	—	—
BYI-A	Beck Youth Inventory Anxiety Subscale (25)	passed	not passed	—	—
CASI-Anx	Child and Adolescent Symptom Inventory (26)	passed	not passed	—	—
EAQ	Existential Anxiety Questionnaire (27)	not passed	—	—	—
FSSC-R	Fear Survey Schedule for Children Revised (28)	not passed	—	—	—
GAD-7	General Anxiety Disorder Assessment (29)	passed	passed	yes	no
HAM-A	Hamilton Anxiety Scale (30)	passed	not passed	—	—
IUSC	Intolerance of Uncertainty Scale for Children (31)	not passed	—	—	—
MASC	Multidimensional Anxiety Scale for Children (32)	passed	not passed	—	—
NEURO-QOL	Neuro QoL Short Form v1.0 Pediatric Anxiety (33)	passed	not passed	—	—
NIH	NIH Toolbox Fear Fixed Form Ages 8 17 v2.0 (34)	not passed	—	—	—
PARS	Paediatric Anxiety Rating Scale (35)	passed	not passed	—	—
PAS	Preschool Anxiety Scale (36)	passed	not passed	—	—
PROMIS	PROMIS Pediatric Short Form v2.0 Anxiety 8a (37)	passed	not passed	—	—
PSWQ-C	Penn State Worry Questionnaire for Children (38)	not passed	—	—	—
RCMAS	Revised Children's Manifest Anxiety Scale (39)	passed	not passed	—	—
SCARED	Screen for Child Anxiety Related Disorders (40)	passed	passed	yes	no
SCAS	Spence Children's Anxiety Scale (41)	passed	passed	yes	no
SEQSS	Self-Efficacy Questionnaire for School Situations (42)	not passed	—	—	—
SFT	School Fear Thermometer (43)	not passed	—	—	—
STAIC	State Trait Anxiety Inventory for Children (44)	passed	not passed	—	—

Table S4. Overview of all Depression Measures Considered

Acronym	Measure Name	Comprehensive Domain Coverage	Feasibility Assessment	Subject to Psychometric Review	Included per Group Consensus
Measures Covering Anxiety & Depression					
CEMS	Children's Emotion Management Scales (15)	not passed	—	—	—
DASS-21	Depression Anxiety Stress Scale (16)	passed	not passed	—	—
HADS	Hospital Anxiety and Depression Scale (17)	passed	not passed	—	—
NASSQ	Negative Affect Self-Statement Questionnaire (18)	not passed	—	—	—
PANAS	Positive and Negative Affect Scale for Children (19)	not passed	—	—	—
RCADS	Revised Child Anxiety and Depression Scales (20)	passed	passed	yes	yes
YSR	Youth Self-Report (21)	passed	not passed	—	—
Measures Covering Depression Only					
ADRS	Adolescent Depression Rating Scale (45)	passed	passed	yes	no
BDI-II	Beck Depression Inventory 2nd version (46)	passed	not passed	—	—
BHS-II	Beck Hopelessness Scale (47)	passed	not passed	—	—
CDI-2	Children's Depression Inventory 2nd version (48)	passed	not passed	—	—
CDRS-R	Children's Depression Rating Scale Revised (49)	passed	not passed	—	—
CES-D	Centre for Epidemiologic Studies Depression Scale (50)	passed	passed	yes	no
DSRSC	Depression Self-Rating Scale for Children (51)	passed	passed	yes	no
HAM-D	Hamilton Rating Scale for Depression (52)	passed	passed	yes	no
MADRS	Montgomery Asberg Depression Rating Scale (53)	passed	passed	yes	no
MFQ	Mood and Feelings Questionnaire (54)	passed	passed	yes	no
NEURO-QOL	Neuro QoL Short Form v1.0 Pediatric Anger (33)	not passed	—	—	—
NEURO-QOL	Neuro QoL Short Form v1.1 Pediatric Depression (33)	passed	not passed	—	—
NIH	NIH Toolbox Anger Affect Fixed Form Ages 8-17 (34)	not passed	—	—	—
NIH	NIH Toolbox Sadness Fixed Form ages 8-17 (34)	not passed	—	—	—
PFC	Preschool Feelings Checklist (55)	not passed	—	—	—
PHQ-9	Patient Health Questionnaire 9 (56)	passed	passed	yes	no
POMS-SF	Profile of Mood States Short Form (57)	not passed	—	—	—
PROMIS	PROMIS Pediatric Scale v2.0 Anger 9a (37)	not passed	—	—	—
PROMIS	PROMIS Pediatric Short Form v2.0 – Depressive Symptoms 8a (37)	passed	passed	yes	no
QIDS-SR	Quick Inventory of Depressive Symptomatology (58)	passed	passed	yes	no
RADS	Reynolds Adolescent Depression Scale (59)	passed	not passed	—	—
TAS-20	Toronto Alexithymia Scale (60)	passed	not passed	—	—

Table S5. Overview of all OCD and PTSD Measures Considered

Acronym	Measure Name	Comprehensive Domain Coverage	Feasibility Assessment	Subject to Psychometric Review	Included per Group Consensus
Symptoms of OCD					
ChOCI-R	Children's Obsessional Compulsive Inventory-Revised (61)	passed	not passed	—	—
CY BOCS	Children's Yale-Brown Obsessive-Compulsive Scale (62)	passed	not passed	—	—
NIMH GOCS	National Institute of Mental Health – Global Obsessive-Compulsive Scale (63)	passed	not passed	—	—
WBSI	White Bear Suppression Inventory (64)	not passed	—	—	—
RBS-R	Repetitive Behaviour Scale – Revised (65)	not passed	—	—	—
ZWIK	Zwangsinventars für Kinder und Jugendliche (66)	passed	not passed	—	—
OCI CV	Obsessive Compulsive Inventory – Revised – Child Version (67)	passed	passed	yes	yes
LOI CV	Leyton Obsessional Inventory – Child Version (68)	passed	passed	yes	no
OBQ CV	Obsessive Beliefs Questionnaire – Child Version (69)	passed	passed	yes	no
Symptoms of PTSD					
CPSS	Child PTSD Symptom Scale (70)	passed	passed	yes	no
UCLA PTSD RI	University of California at Los Angeles PTSD Reaction Index (71)	passed	not passed	—	—
CAPS CA	Clinician Administered PTSD Scale for Children and Adolescents (72)	passed	not passed	—	—
CRIES	Children's Revised Impact of Event Scale (73)	passed	passed	yes	yes
TSCC	Trauma Symptom Checklist for Children (74)	passed	not passed	—	—
PDS	Posttraumatic Diagnostic Scale (75)	passed	not passed	—	—

Table S6. Overview of all Functioning and Impairment Measures Considered

Acronym	Measure Name	Comprehensive Domain Coverage	Feasibility Assessment	Subject to Psychometric Review	Included per Group Consensus
Global Functioning					
BIS	Brief Impairment Scale (76)	passed	passed	yes	no
BPRS	Brief Psychiatric Rating Scale (77)	passed	passed	yes	no
CAFAS	Child and Adolescent Functional Assessment (78)	passed	not passed	—	—
CGAS	Children's Global Assessment Scale (79)	passed	passed	yes	yes
CGI-I	Clinical Global Impressions Scale – Improvement (80)	passed	passed	yes	no
CGI-S	Clinical Global Impressions Scale – Severity (80)	passed	passed	yes	no
CIS	Columbia Impairment Scale (81)	passed	not passed	—	—
CSDS	Child Sheehan Disability Scale (82)	passed	not passed	—	—
HoNOSCA	Health of the Nation Outcome Scale (83)	passed	passed	yes	no
IRS	Impairment Rating Scale (84)	passed	not passed	—	—
ORS	Outcome Rating Scale (85)	passed	not passed	—	—
SDQ	Strengths and Difficulties Questionnaire Impact Supplement (86)	passed	not passed	—	—
SLOF	Specific Levels of Functioning Scale (87)	passed	not passed	—	—
Vineland-II	Vineland Adaptive Behavior Scales–Second Edition (88)	passed	not passed	—	—

Table S7. Overview of all HRQoL Measures Considered

Acronym	Measure Name	Comprehensive Domain Coverage	Feasibility Assessment	Maximum Age Range Covered	Fewer Than 20 Items [a]	Subject to Psychometric Review [a]	Included per Group Consensus
Disorder-Specific Impairment							
CAIS	Child's Anxiety Impact Scale (89)	passed	not passed	—	no	—	—
CALIS	Children's Anxiety Life Interference Scale (90)	passed	passed	passed	passed	yes	yes
COIS-R	Child OC Impact Scale – Revised (91)	not passed	—	—	—	—	—
Health-Related Quality of Life (HRQoL)							
CHQ	Child Health Questionnaire (92)	passed	not passed	—	—	—	—
CHU-9D	Child Health Utility 9D (93)	passed	not passed	—	—	—	—
EQ-5D-Y	EuroQol 5D-youth (94)	passed	passed	passed	passed	yes	no
KIDSCREEN	KIDSCREEN 10, 27,52 (95)	passed	passed	passed	passed	yes	yes
Kindl-r	Kindl-r 24 (96)	passed	passed	passed	not passed	—	—
PEDSQL	Pediatric Quality of Life Inventory (97)	passed	not passed	—	—	—	—
PQ-LES-Q	Pediatric Quality of Life and Enjoyment Satisfaction Questionnaire (98)	passed	passed	passed	passed	yes	no
PROMIS	PROMIS Pediatric Global Health (99)	passed	passed	passed	passed	yes	no
SF-12/SF-36	Medical Outcomes Study 12 or 36-item Short Form Health Survey (100)	passed	not passed	—	—	—	—
SIP	Sickness Impact Profile (101)	passed	not passed	—	—	—	—
WHODAS Child	World Health Organization Disability Assessment Schedule for Children (102)	passed	passed	passed	not passed	—	—
WHOQOL	World Health Organisation Quality of Life Assessment (103)	passed	passed	not passed	—	—	—
YQOL-R	Youth Quality of Life Instrument – Research (104)	passed	not passed	—	—	—	—

[a] These additional criteria were applied to the appraisal of impairment and HRQoL measures as part of adapting the process to the symptom and functioning measures that had already been selected when this last group of measures were reviewed.

Identification and assessment of case-mix factors. Drawing on the initial systematic review and existing ICHOM Standard Sets, the central project team (KK, SC, MW) compiled a list of candidate variables to record for the purpose of case-mix adjustment. One member of the project team (KK) conducted a rapid review of reviews considering seven systematic reviews and meta-analyses to examine the evidence base on the role of each possible case-mix factor as a predictor or moderator of treatment response. Findings were presented to the Working Group, who then appraised each case mix factor for its relevance (i.e., strength of the evidence for a causal linkage with treatment response), and for the feasibility of collecting and comparing it practically and cross-nationally. Table S8 provides an overview of the reviews considered, and tables S9 and S10 provide a summary of the review findings in relation to different case mix factors.

Table S8. Overview of Reviews Considered for the Rapid Review of Reviews (Case-Mix Factors)

	Disorder(s) covered	Type of Review	No. of studies reviewed (*K*)	Case mix factors covered
Emslie et al. 2011 (105)	Depression	Review	*K* = 3	Age; gender; ethnic minority status; socio-economic status; baseline symptom severity; symptom duration; presence of comorbidities; maltreatment
Knight et al. 2014 (106)	Anxiety	Systematic Review	*K* = 51	Age; gender; ethnic minority status; socio-economic status; baseline symptom severity; presence of comorbidities; parental mental health
Lundkvist-Houndoumadi et al., 2014 (107)	Anxiety	Systematic Review	*K* = 24	Age; gender; ethnic minority status; socio-economic status; baseline symptom severity; presence of comorbidities; parental mental health
Nanni et al. 2012 (108)	Depression	Meta-analysis	*K* = 10	Maltreatment
Nilsen et al., 2013 (109)	Anxiety & Depression	Systematic Review	*K* = 32	Age; gender; ethnic minority status; baseline symptom severity; symptom; presence of comorbidities
Skriner et al. 2019 (110)	Anxiety	Integrative analysis	*K* = 9	Age; gender; ethnic minority status; presence of comorbidities
Walczak et al. 2018 (111)	Anxiety	Systematic Review	*K* = 25	Presence of comorbidities

Table S9. Rapid Review of Reviews on Predictors and Moderators of Treatment Response for Anxiety

	Number of individual studies out of all relevant studies assessed by the review that showed significant associations between the case mix variable and treatment response				
	Nilsen et al., 2013	Lundkvist-Houndoumadi et al., 2014	Knight et al. 2014	Walczak et al. 2018	Skriner et al. 2019 [a]
Demographic factors					
Age	5/21	3/9	2/11	—	Yes [b]
Gender	4/21	2/10	1/10	—	Yes [b]
Ethnic minority status	1/6	0/3	0/8	—	No [c]
Socio-economic status	—	0/4	0/10	—	—
Household situation	—	—	—	—	—
School attendance	—	—	—	—	—
Clinical factors					
Baseline symptom severity	2 / 6	7/8	3/10	—	—
Duration of symptoms	1 / 2	—	—	—	—
Presence of comorbidities	4/17	3/9	7/23	16/25	Yes [b]
Parental mental health	—	7/10	6/11	—	—
Maltreatment	—	—	—	—	—

[a] This review did not report the number of reviewed studies observing significant effects, but presented meta-analysis results.
[b] Meta-analysis showed a significant association with treatment response.
[c] Meta-analysis showed no significant association with treatment response.

Table S10. Rapid Review of Reviews on Predictors and Moderators of Treatment Response for Depression

	Number of individual studies out of all the studies assessed by the relevant review that showed significant associations between the case mix variable and treatment response		
	Emslie et al. 2011	Nanni et al. 2012 [a]	Nilsen et al. 2013
Demographic factors			
Age	1/3	—	2/5
Gender	0/3	—	0/7
Ethnic minority status	0/3	—	2/3
Socio-economic status	1/1	—	—
Household situation	—	—	—
School attendance	—	—	—
Clinical factors			
Baseline symptom severity	3/3	—	4/4
Duration of symptoms	2/3	—	0/2
Presence of comorbidities	3/3	—	3/6
Parental mental health	—	—	—
Maltreatment	0/2	Yes [b]	—

[a] This review did not report the number of reviewed studies observing significant effects, but presented meta-analysis results.
[b] Meta-analysis showed no significant association with treatment response.

Open Review of the Draft Recommendations. To obtain external feedback from the wider practitioner and service user communities on the acceptability of the Working Group's recommendations, the draft Standard Set was compiled into an anonymous online survey along with feedback questions and additional free text response options. This *Open Review* survey was disseminated to mental health professionals, researchers, individuals with lived experience of service use as a child or young person, and their parents through the same channels used for the recruitment of Working Group members. This included professional organisations, service user associations, the networks of ICHOM and the project funders, social media, and the Working Group's networks. The survey targeting professionals was accessible globally and yielded 463 responses from 45 countries (table S11). The survey targeting individuals and parents engaged participants in Denmark, the United States (US), and the United Kingdom (UK) due to varied national requirements for ethical approval. This yielded responses from 24 individuals, of which 9 indicated experience as a parent or carer. Open Review participants provided over 200 free-text comments on the proposed measures. The project team (KK & SC) first categorised these comments using framework analysis (112), and subsequently screened each comment for instrument-specific sub-themes (e.g., item content, feasibility, psychometric properties) using inductive thematic analysis (113). Common themes were reviewed and discussed by the Working Group. The Open Review survey was anonymous and open to any interested member of the public over 18 years of age.

Table S11. Composition of the Practitioner Sample in the Open Review

Region / Subregion	***n* (%)**
Africa	17 (3.7%)
Americas	57 (12.3%)
North America	46 (9.9%)
Latin America & the Caribbean	11 (2.4%)
Asia	44 (9.6%)
Eastern and Southern Asia	22 (4.8%)
Western Asia	22 (4.8%)
Europe	305 (65.8%)
Northern Europe	258 (55.7%)
Western Europe	26 (5.6%)
Southern Europe	14 (3.0%)
Eastern Europe	7 (1.5%)
Oceania	40 (8.6%)
Total	467 (100.0%)

Supplementary Information on the Selected Outcome Domains and Measurement Instruments

From the list of 70 candidate outcomes, the Working Group selected nine: *symptoms of anxiety*, *symptoms of depression*, *symptoms of OCD*, *symptoms of PTSD*, *suicidal thoughts and behaviour*, *global functioning*, *sleep functioning, social functioning*, and *peer relationships*. These were grouped into the three high-level domains of *symptoms*, *suicidal thoughts and behaviour*, and *functioning*. The Group agreed that symptoms of anxiety, depression, OCD, and PTSD; suicidal thoughts and behaviour; and global functioning should be measured through dedicated instruments, and reported separately. In turn, it was decided that sleep functioning, social functioning, and peer relationships would not require separate measurement via dedicated instruments if they could be adequately covered by a composite measure(s) of global functioning or symptoms. This decision was taken to minimise the complexity and duplication within the Standard Set, enhancing feasibility and uptake.

Following the systematic appraisal of relevant measurement instruments for the selected outcome domains, the Working Group selected seven measures. The Revised Children's Anxiety and Depression Scale Short Version (RCADS-25) was selected as a joint measure of anxiety and depression symptoms, which allows for the computation of separate anxiety and depression total scores (114). The Obsessive Compulsive Inventory for Children (OCI-CV) was selected as a measure of OCD symptoms (67), the Children's Revised Impact of Events Scale (CRIES) as a measure of PTSD symptoms (73), and the Columbia Suicide Severity Rating Scale (C-SSRS) as a measure of suicidal thoughts and behaviour (115). The Working Group struggled to identify a clinician- or self-reported measure of functioning that satisfied all selection criteria. Common issues included overlap in item coverage between symptoms and functioning, a perceived overemphasis on bodily functions as opposed to psychosocial functioning, lack of validation across the full age span or in clinical populations, and cross-cultural validity.

In the absence of a single measure that fully satisfied all criteria, the group formed consensus on tracking both a broad concept of global functioning or health-related quality of life (HRQoL), and condition-specific functional impairment, through dedicated measures of each concept. The Group selected the self-reported KIDSCREEN-10 (116) and the clinician-rated Children's Global Assessment Scale (CGAS; 117) as measures of global functioning, and the self-reported Children's Anxiety Life Interference Scale (CALIS; 90) as a measure of condition-specific impairment. Each scale is described in more detail below and in table 1 of the manuscript.

All measurement instruments included in the Standard Set can be completed in less than 20 minutes, are freely available for use in clinical practice (table S12), have been validated for use with CYP, and have been translated into at least one additional language. The Working Group only selected instruments that are suitable for paper-, as well as electronic-based administration, to ensure their feasibility in service settings without well-developed digital systems.

Symptoms of anxiety and depression. The Revised Children's Anxiety and Depression Scale Short Version (RCADS-25) is a self- and parent-report scale for youth aged 8-18 years, measuring the frequency of symptoms associated with depression and anxiety (114). It is a shortened version of the original RCADS-47 (20), which consists of six subscales assessing symptoms of major depressive disorder (MDD, 10 items), generalized anxiety disorder (GAD, 6 items), separation anxiety disorder (SAD, 7 items), social phobia (SP, 9 items), panic disorder (PD, 9 items), and obsessive-compulsive disorder (OCD, 6 items), in line with DSM-IV dimensions. Symptom frequency is scored on a four-point Likert scale ranging from 0 (*never*) to 3 (*always*).

The RCADS-25 was derived from the RCADS-47 through bifactor-modelling (114). It includes three items each to cover GAD, OCD, PD, SAD, and SP. While it does not allow for the computation of separate subscales for each of these anxiety disorders, the 15 items can be aggregated into an overall anxiety score. The RCADS-25 further includes the same 10 items covering major depressive disorder as the original long form. The RCADS-25 is currently available in English, Hindi, Spanish and Swedish. The original 47-item version is available in 14 languages. Long or short versions of the RCADS have been applied in at least 25 countries across Africa, Europe, North America, South America, and Asia (118). The RCADS-47 has demonstrated good internal consistency, reliability, and construct validity (20,119–122).There is also some evidence of its sensitivity to change. While the RCADS-25 has been less widely validated, it has demonstrated high internal consistency for its self- and parent-reported anxiety subscale (Cronbach's $\alpha = 0.82 – 0.95$), depression subscale ($\alpha = .79 – .93$), and total scale ($\alpha = .88 – .90$), and satisfactory test-retest reliability ($r = .70 – .90$) for the subscales and total scale in clinical and non-clinical samples (114,123–125). Additional research is needed to examine its sensitivity to change, and structural validity (125).

Symptoms of OCD. The 21-item OCI-CV assesses OCD symptom severity across six symptom domains: doubting/checking obsessing; hoarding; washing; ordering; and neutralising. Symptom frequency is scored in a 3-point Likert scale ranging from 0 (*never*) to 2 (*always*). The OCI-CV has been validated in clinical and non-clinical populations in Africa, Asia, Europe, North and South America. It has demonstrated good internal consistency (Cronbach's α = .83 – .91) and acceptable test-retest reliability (*r* = 0.77 – 0.82; *ICC* = .85 – .92) in clinical and non-clinical populations (67,126–130), with some evidence for its sensitivity to change (67). A parent-version of the OCI-CV is not currently available, and the Standard Set recommends collecting only youth report until a matching parent-report becomes available. The OCI-CV aligns with the ICHOM adult Standard Set for anxiety and depression, which recommends its adult version (6).

Symptoms of PTSD. The Standard Set includes the Children's Revised Impact of Events Scale's 8-item version (CRIES-8) for youth self-report, and its 13-item version for parent report (CRIES-13; 73). The CRIES-8 measures intrusions and avoidance; the CRIES-13 includes five additional items measuring hyperarousal. Symptom frequency is scored on a four-point scale (0, *not at all;* 1, *rarely*; 3, *sometimes*; and 5, *often*). The CRIES has been used in research in Europe, East and South Asia, and Asia Pacific, and is available in at least 27 languages. The CRIES-8 and CRIES-13 have demonstrated satisfactory test-retest reliability (*r* = 0.78 and 0.85, respectively) in a clinical sample (131), high internal consistency (Cronbach's α = .70 – .86 for the CRIES-8; and .74 – .89 for the CRIES-13), in clinical and non-clinical samples (131–133). The parent version has shown evidence of sensitivity to change (134).

Suicidal thoughts and behaviour. The Columbia Suicide Severity Rating Scale (C-SSRS) Recent Self-Report Screener (115) is a shorter self-report version of the clinician-administered semi-structured C-SSRS interview protocol. The C-SSRS Recent Self-Report Screener consists of five items assessing the severity of suicidal ideation in the previous month (i.e., suicidal thoughts, intent, and plans), and a sixth item that assesses suicidal behaviour. Items are scored on a binary response scale response scale (*yes*/*no*). While the self-report screener has not yet been validated in CYP, the severity of ideation subscale of the clinician-rated C-SSRS (from which most of the Self-Report Screener is derived) has demonstrated good internal consistency (ordinal α = .97) and interrater reliability(*IRR* = .92), as well as sensitivity to change in clinical adolescent samples (115). The C-SSRS is available in at least 100 languages.

Global functioning. The KIDSCREEN-10 is part of the KIDSCREEN suite of measures, which assesses Health-Related Quality of Life (HRQoL) in CYP aged 8-18 years, and also includes the more comprehensive KIDSCREEN-52 and KIDSCREEN-27 (135). The KIDSCREEN-52 was initially developed through a process of cross-cultural harmonisation across 13 European countries. The shorter KIDSCREEN-27 was developed subsequently, and consists of 27 items across 5 dimensions of HRQoL. The KIDSCREEN-10 was derived from the 27-item version using Rasch analysis (51) in order to offer an even briefer assessment that could be used in large population-based studies or for routine monitoring purposes without imposing an undue burden on respondents and administrators (116). Responses are scored on a five-point Likert scale that indicates the frequency or intensity of specific experiences (from *never/not at all* to *always / extremely*). The KIDSCREEN-10 has shown satisfactory internal consistency for the youth- (Cronbach α = 0.80 – 0.82) and parent-version (Cronbach α = 0.76 – 0.78), and satisfactory test-retest reliability for the youth version (ICC = 0.70) (116,136,137).

The Children's Global Assessment Scale (CGAS) is a brief clinician-rated measure of global functioning (117). The CGAS was developed for CYP aged 4-16 years, and has been widely used over the past three decades. Clinicians preform a single rating by locating CYP on a scale from 1 to 100, which places them into one of ten categories, from 1 – 10 (*extremely impaired*) to 91 – 100 (*doing very well*). The CGAS has demonstrated good inter-rater and test-retest reliability (117,138), and sensitivity to change (139).

Condition-specific impairment. Condition-specific impairment relates to the "ways in which symptoms interfere with and reduce adequate performance of important and desired aspects of a child's life" (14). The Standard Set recommends the Children's Anxiety Life Interference Scale (CALIS; 90) to measure this aspect of functioning. The CALIS consists of 9 items (10 in the parent version), which describe instances where anxiety symptoms may impact on functioning at home, at school, on social life, or on activities. The extent of symptom interference with daily functioning is scored on a five-point Likert scale from 0 (*not at all*) to 4 (*a great deal*). The CALIS was developed for CYP aged 6-17 years and is available in at least seven languages. Both the youth and parent versions have demonstrated high internal consistency (Cronbach's α = 0.71 – 0.90) and largely acceptable test-retest reliability (*r* = 0.66 – 0.87) in clinical and non-clinical samples, as well as evidence for sensitivity to change (90,140,141).

Casemix factors. The Standard Set recommends using the Current View Tool (142) to measure comorbid presenting problems, and complexity factors at baseline, for the purpose case mix adjustment. This brief clinician-rated screening tool is completed by clinicians based on all available information about a young person and their family, including other completed standardised measures. It includes a battery of 30 brief problem descriptions that map onto ICD-11 diagnostic criteria considered relevant to CYP (2). Each presenting problem is scored according to the severity of distress and functional impairment it is causing, using a scale from 0 (*none*) to 3 (*severe*). The Current View Tool further allows clinicians to indicate the presence of 14 complexity factors. These include issues related to the young person's health, such as the presence of serious physical health issues, a pervasive developmental disorder, or neurological disorders; issues related to the family situation, such as whether a young person has been taken into the care of their local authority, whether they have caring duties for a family member, whether their wellbeing is being monitored by social services, whether parents experience health issues, and whether the family is known to live in financial difficulty. Other complexity factors include being a refugee or asylum seeker; having experienced war, torture or trafficking; having experienced abuse or neglect; or having been in contact with the youth justice system. In addition, clinicians can mark whether young people have difficulties with attendance and/or attainment at school, training, or work. The Current View Tool was developed in the United Kingdom, and may require adjustment for the purpose of cross-cultural validity. It is hoped that insights from the piloting of the Standard Set will be able to inform such efforts in the future. The Working Group considered, for example, that it would be important to add exposure to natural disaster as a complexity factor, which is not currently covered by the Current View Tool.

Table S12. Information on Licencing and Access to the Recommended Measurement Instruments

Concept	Questionnaire	Informant(s)	Reference	Licensing information and web link
Symptoms of anxiety & depression	Revised Children's Anxiety and Depression Scale Short version (RCADS-25)	CYP & parents	Ebesutani et al., 2010, 2012 (114,143)	Copyrighted by Susan Spence and Bruce Chorpita. No fee or license required for use. Please notify Dr. Chorpita before use in published studies and read and understand the terms of use available via the link. Questionnaire can be accessed here: https://www.childfirst.ucla.edu/resources/
Symptoms of OCD	Obsessive Compulsive Inventory for Children (OCI-CV)	CYP	Foa et al., 2010 (67)	Copyrighted by Edna Foa. No fee or license required for use in healthcare settings. Requests for access to be sent to Ellen Kubis (ekubis@pennmedicine.upenn.edu)
Symptoms of PTSD	Children's Revised Impact of Events Scale (CRIES-8/13)	CYP & parents	Yule 1997 (73)	Copyrighted by the Children and War Foundation. No fee or license required for use. Questionnaire can be accessed here: https://www.childrenandwar.org/projectsresources/measures/
Suicidal thoughts and behaviour	Columbia Suicide Severity Rating Scale (C-SSRS) Recent Self-Report Screener	CYP	Posner et al., 2011 (115)	Copyrighted by Research Foundation for Mental Hygiene. No fee or license required for use in community and healthcare settings, or in non-profit research. For inquiries and training requirements contact posnerk@nyspi.columbia.edu. Questionnaire can be accessed here: http://cssrs.columbia.edu/the-columbia-scale-c-ssrs/about-the-scale/
Global functioning	KIDSCREEN-10	CYP & parents	Ravens-Sieberer et al., 2010 (116)	Copyrighted by the KIDSCREEN group. No fee for funded and non-funded academic research and non-commercial organisation research and evaluation studies. Fees required for commercial usage. Collaboration form required. Questionnaire can be accessed here: https://www.kidscreen.org/english/questionnaires/kidscreen-10-index/
	Children's Global Assessment Scale (CGAS)	Clinician	Shaffer 1983, (117)	No fee or license required for use. Questionnaire can be accessed here: https://www.cymh.ca/modules/MeasuresDatabase/en/Home/Detail/77#Key-Info-Content
Impact of condition on daily life	Children's Anxiety Life Interference Scale (CALIS)	CYP & parents	Lyneham et al., 2013 (90)	Copyrighted by Centre for Emotional Health, Macquarie University, Sydney, Australia. No fee or license required for research or clinical purposes. Questionnaire can be accessed here: https://www.mq.edu.au/research/research-centres-groups-and-facilities/healthy-people/centres/centre-for-emotional-health-ceh/resources
Casemix: Presenting problems & complexity factors	Current View Tool	Clinician	Jones et al., 2013 (142)	Copyrighted. No fee or license required. Questionnaire can be accessed here: https://www.corc.uk.net/outcome-experience-measures/current-view/

References

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition. Arlington, VA: American Psychiatric Association; 2013.
2. World Health Organization. International Classification of Diseases for Mortality and Morbidity Statistics (11th Revision) [Internet]. 2018. Available from: https://icd.who.int/browse11/l-m/en [15/10/2019].
3. Regier DA, Kuhl EA, Kupfer DJ. The DSM-5: Classification and criteria changes. World Psychiatry. 2013;12(2):92–8.
4. Garber J, Weersing VR. Comorbidity of Anxiety and Depression in Youth: Implications for Treatment and Prevention. Clin Psychol Sci Pract. 2010;17(4):293–306.
5. Cummings CM, Caporino NE, Kendall PC. Comorbidity of anxiety and depression in children and adolescents: 20 years after. Psychol Bull. 2014;140(3):816–45.
6. Obbarius A, van Maasakkers L, Baer L, Clark DM, Crocker AG, de Beurs E, et al. Standardization of health outcomes assessment for depression and anxiety: recommendations from the ICHOM Depression and Anxiety Working Group. Qual Life Res. 2017;26(12):3211–25.
7. Biernacki P, Waldorf D. Snowball Sampling: Problems and Techniques of Chain Referral Sampling. Sociol Methods Res. 1981. 10(2):141–63.
8. Fleiss JL. Statistical methods for rates and proportions. New York, NY: John Wiley and Sons; 1981. 218 p.
9. Krause KR, Bear HA, Edbrooke-Childs J, Wolpert M. Review: What Outcomes Count? A Review of Outcomes Measured for Adolescent Depression Between 2007 and 2017. J Am Acad Child Adolesc Psychiatry. 2019;58(1):61–71.
10. Hoagwood KE, Jensen PS, Petti T, Burns BJ. Outcomes of mental health care for children and adolescents: I. A comprehensive conceptual model. J Am Acad Child Adolesc Psychiatry. 1996;35(8):1055–63.
11. World Health Organization. International classification of functioning, disability and health: ICF [Internet]. World Report on Child Injury Prevention. Geneva; 2008. Available from: https://apps.who.int/iris/bitstream/handle/10665/43851/9789241563574_eng.pdf;jsessionid=39FF8A78C7D7F436BCE40AE9AC872A51?sequence=1 [20/06/2020].
12. Reeve BB, Wyrwich KW, Wu AW, Velikova G, Terwee CB, Snyder CF, et al. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. Qual Life Res. 2013;22(8):1889–905.
13. Terwee CB, Jansma EP, Riphagen II, De Vet HCW. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. Qual Life Res. 2009;18(8):1115–23.
14. Rapee RM, Bögels SM, Van Der Sluis CM, Craske MG, Ollendick T. Annual research review: Conceptualising functional impairment in children and adolescents. J Child Psychol Psychiatry Allied Discip. 2012;53(5):454–68.
15. Zeman J, Shipman K, Suveg C. Anger and Sadness Regulation: Predictions to Internalizing and Externalizing Symptoms in Children. J Clin Child Adolesc Psychol. 2002;31(3):393–8.
16. 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.
17. Zigmond AS, Snaith RP. The Hospital Anxiety and Depression Scale. Acta Psychiatr Scand. 1983 Jun;67(6):361–70.
18. Ronan KR, Kendall PC, Rowe M. Negative affectivity in children: Development and validation of a self-statement questionnaire. Cognit Ther Res. 1994;18(6):509–28.
19. Hughes AA, Kendall PC. Psychometric properties of the positive and negative affect scale for children (PANAS-C) in children with anxiety disorders. Child Psychiatry Hum Dev. 2009 Sep;40(3):343–52.
20. Chorpita BF, Yim L, Moffitt C, Umemoto LA, Francis SE. Assessment of symptoms of DSM-IV anxiety and depression in children: A revised child anxiety and depression scale. Behav Res Ther. 2000;38(8):835–55.
21. Achenbach TM. Integrative Guide to the 1991 CBCL/4–18, YSR, and TRF Profiles. Burlington, VT: University of Vermont, Department of Psychology; 1991.
22. Silverman WK, Albano AM. Anxiety Disorders Interview Schedule (Adis-IV) Child Interview Schedule. Vol. 2. Graywind Publications; 2006.
23. Steer RA, Beck AT. Beck Anxiety Inventory. In: Zalaquett CP, Wood RJ, editors. Evaluating stress: A book of resources. Scarecrow Education; 1997. p. 23–40.
24. Carleton RN, McCreary DR, Norton PJ, Asmundson GJG. Brief fear of negative evaluation scale - Revised.

Depress Anxiety. 2006;23(5):297–303.
25. Beck JS, Beck AT, Jolly JB, Steer RA. Beck Youth Inventories-Second Edition for Children and Adolescents manual. San Antonio, TX: PsychCorp; 2005.
26. Sukhodolsky DG, Scahill L, Gadow KD, Arnold LE, Aman MG, McDougle CJ, et al. Parent-rated anxiety symptoms in children with pervasive developmental disorders: Frequency and association with core autism symptoms and cognitive functioning. J Abnorm Child Psychol [Internet]. 2008;36(1):117–28.
28. Ollendick TH. Reliability and validity of the revised fear survey schedule for children (FSSC-R). Behav Res Ther. 1983;21(6):685–92.
29. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: The GAD-7. Arch Intern Med. 2006;166(10):1092–7.
30. Maier W, Buller R, Philipp M, Heuser I. The Hamilton Anxiety Scale: reliability, validity and sensitivity to change in anxiety and depressive disorders. J Affect Disord. 1988;14(1):61–8.
31. Comer JS, Roy AK, Furr JM, Gotimer K, Beidas RS, Dugas MJ, et al. The Intolerance of Uncertainty Scale for Children: A Psychometric Evaluation. Psychol Assess. 2009 Sep;21(3):402–11.
32. March JS, Parker JDA, Sullivan K, Stallings P, Conners CK. The Multidimensional Anxiety Scale for Children (MASC): Factor structure, reliability, and validity. J Am Acad Child Adolesc Psychiatry. 1997 Apr;36(4):554–65.
33. Lai J-S, Nowinski C, Victorson D, Bode R, Podrabsky T, McKinney N, et al. Quality-of-Life Measures in Children With Neurological Conditions. Neurorehabil Neural Repair. 2012;26(1):36–47.
34. Salsman JM, Butt Z, Pilkonis PA, Cyranowski JM, Zill N, Hendrie HC, et al. Emotion assessment using the NIH Toolbox. Neurology. 2013;80(11 Supplement 3):S76–86.
35. Research Units on Pediatric Psychopharmacology Anxiety Study Group. The Pediatric Anxiety Rating Scale (PARS): Development and Psychometric Properties. J Am Acad Child Adolesc Psychiatry. 2002 Sep;41(9):1061–9.
36. Edwards SL, Rapee RM, Kennedy SJ, Spence SH. The assessment of anxiety symptoms in preschool-aged children: The revised preschool anxiety scale. J Clin Child Adolesc Psychol. 2010;39(3):400–9.
37. Varni JW, Magnus B, Stucky BD, Liu Y, Quinn H, Thissen D, et al. Psychometric properties of the PROMIS® pediatric scales: Precision, stability, and comparison of different scoring and administration options. Qual Life Res. 2014;23(4):1233–43.
38. Meyer TJ, Miller ML, Metzger RL, Borkovec TD. Development and validation of the penn state worry questionnaire. Behav Res Ther. 1990;28(6):487–95.
39. Reynolds CR, Richmond BO. What I think and feel: A revised measure of children's manifest anxiety. J Abnorm Child Psychol. 1978;6(2):271–80.
40. Birmaher B, Khetarpal S, Brent D, Cully M, Balach L, Kaufman J, et al. The Screen for Child Anxiety Related Emotional Disorders (SCARED): Scale construction and psychometric characteristics. J Am Acad Child Adolesc Psychiatry. 1997 Apr;36(4):545–53.
41. Spence SH. A measure of anxiety symptoms among children. Behav Res Ther. 1998;36(5):545–66.
42. Heyne D, King N, Tonge B, Rollings S, Pritchard M, Young D, et al. The self-efficacy questionnaire for school situations: Development and psychometric evaluation. Behav Chang. 1998;15(1):31–40.
43. Heyne D, Rollings S, King NJ, Tonge BJ. School refusal. Oxford, England: Wiley-Blackwell; 2002.
44. Spielberger C. Manual for the State-Trait Anxiety Inventory for Children. Palo Alto, CA: Consulting Psychologists Press; 1973.
45. Revah-Levy A, Birmaher B, Gasquet I, Falissard B. The Adolescent Depression Rating Scale (ADRS): A validation study. BMC Psychiatry. 2007;7:2.
46. Beck AT, Steer RA, Carbin MG. Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. Clin Psychol Rev. 1988;8(1):77–100.
47. Beck A, Steer R. Beck hopelessness scale (BHS) manual. San Antonio, TX: Pearson; 1993.
48. Kovacs M. Rating scales to assess depression in school-aged children. Acta Paedopsychiatr. 1981;46(5–6):305–15.
49. Poznanski E, Mokros H. Children's Depression Rating Scale–Revised (CDRS-R). Los Angeles, CA: Western Psychological Services; 1996.
50. Radloff LS. The CES-D Scale. Appl Psychol Meas. 1977;1(3):385–401.
51. Birleson P. the Validity of Depressive Disorder in Childhood and the Development of a Self-Rating Scale: a Research Report. J Child Psychol Psychiatry. 1981;22(1):73–88.
52. Hamilton M. The Hamilton rating scale for depression. In: Sartorius N, Ban, Thomas A, editors. Assessment of Depression. Berlin, Germany: Springer; 1986. p. 143–52.

53. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. Br J Psychiatry. 1979;134(4):382–9.
54. Angold A, Costello EJ, Messer SC, Pickles A. Development of a short questionnaire for use in epidemiological studies of depression in children and adolescents. Int J Methods Psychiatr Res. 1995;5(4):237–49.
55. Luby JL, Heffelfinger A, Koenig-McNaught AL, Brown K, Spitznagel E. The preschool feelings checklist: A brief and sensitive screening measure for depression in young children. J Am Acad Child Adolesc Psychiatry. 2004 Jun;43(6):708–17.
56. Spitzer RL, Kroenke K, Williams JBW. Validation and utility of a self-report version of PRIME-MD: The PHQ Primary Care Study. J Am Med Assoc. 1999;282(18):1737–44.
57. Dilorenzo TA, Bovbjerg DH, Montgomery GH, Valdimarsdottir H, Jacobsen PB. The application of a shortened version of the profile of mood states in a sample of breast cancer chemotherapy patients. Br J Health Psychol. 1999;4(4):315–25.
58. Rush AJ, Trivedi MH, Ibrahim HM, Carmody TJ, Arnow B, Klein DN, et al. The 16-item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): A psychometric evaluation in patients with chronic major depression. Biol Psychiatry. 2003 Sep;54(5):573–83.
59. Reynolds W. Reynolds Adolescent Depression Scale. Professional manual. Odessa, FL: Psychological Assessment Resources, Inc; 1987.
60. Bagby RM, Parker JDA, Taylor GJ. The twenty-item Toronto Alexithymia scale-I. Item selection and cross-validation of the factor structure. J Psychosom Res. 1994;38(1):23–32.
61. Shafran R, Frampton I, Heyman I, Reynolds M, Teachman B, Rachman S. The preliminary development of a new self-report measure for OCD in young people. J Adolesc. 2003;26(1):137–42.
62. Scahill L, Riddle MA, McSwiggin-Hardin M, Ort SI, King RA, Goodman WK, et al. Children's Yale-Brown Obsessive Compulsive Scale: reliability and validity. J Am Acad Child Adolesc Psychiatry. 1997 Jun;36(6):844–52.
63. Taylor S. Assessment of obsessions and compulsions: Reliability, validity, and sensitivity to treatment effects. Clin Psychol Rev. 1995;15(4):261–96.
64. Muris P, Merckelbach H, Horselenberg R. Individual differences in thought suppression. The White Bear Suppression Inventory: Factor structure, reliability, validity and correlates. Behav Res Ther. 1996;34(5–6):501–13.
65. Mirenda P, Smith IM, Vaillancourt T, Georgiades S, Duku E, Szatmari P, et al. Validating the repetitive behavior scale-revised in young children with autism spectrum disorder. J Autism Dev Disord. 2010;40(12):1521–30.
66. Goletz H, Döpfner M. ZWIK, Zwangsinventar für Kinder und Jugendliche. In: Barkman C, Schule-Markwort M, Brähler E, editors. Klinisch-Psychiatrische Ratingskalen Für Das Kindes-Und Jugendalter. Göttingen, DE: Hogrefe; 2011. p. 489–93.
67. Foa EB, Coles M, Huppert JD, Pasupuleti R V., Franklin ME, March J. Development and Validation of a Child Version of the Obsessive Compulsive Inventory. Behav Ther. 2010;41(1):121–32.
68. Berg CJ, Rapoport JL, Flament M. The Leyton obsessional inventory-child version. J Am Acad Child Psychiatry. 1986;25(1):84–91.
69. Coles ME, Wolters LH, Sochting I, de Haan E, Pietrefesa AS, Whiteside SP. Development and initial validation of the obsessive belief questionnaire-child version (OBQ-CV). Depress Anxiety. 2010;27(10):982–91.
70. Foa EB, Johnson KM, Feeny NC, Treadwell KRH. The Child PTSD Symptom Scale: A Preliminary Examination of its Psychometric Properties. J Clin Child Adolesc Psychol. 2001 Sep;30(3):376–84.
71. Steinberg AM, Brymer MJ, Decker KB, Pynoos RS. The University of California at Los Angeles Post-traumatic Stress Disorder Reaction Index. Curr Psychiatry Rep. 2004;6(2):96–100.
72. Newman E, Weathers FW, Nader K, Kaloupek DG, Pynoos RS, Blake DD, et al. Clinician-Administered PTSD Scale for Children and Adolescents (CAPSCA) Interviewer's Guide. Los Angeles, CA: Western Psychological Services; 2004.
73. Yule W. Anxiety, Depression and Post-Traumatic Stress in Childhood. In: Sclare I, editor. Child Psychology Portfolio. Windsor, England: NFER-Nelson; 1997.
74. Briere J. Trauma Symptom Checklist for Children (TSCC), Professional Manual. 1996. Odessa, FL: Psychological Assessment Resources;
75. Foa EB, Cashman L, Jaycox L, Perry K. The validation of a self-report measure of posttraumatic stress disorder: The posttraumatic diagnostic scale. Psychol Assess. 1997;9(4):445–51.
76. Bird HR, Canino GJ, Davies M, Ramírez R, Chávez L, Duarte C, et al. The Brief Impairment Scale (BIS): A

multidimensional scale of functional impairment for children and adolescents. J Am Acad Child Adolesc Psychiatry. 2005;44(7):699–707.
77. Overall JE, Gorham DR. The Brief Psychiatric Rating Scale. Psychol Rep. 1962;10(3):799–812.
78. Hodges K, Wong MM. Psychometric characteristics of a multidimensional measure to assess impairment: The child and adolescent functional assessment scale. J Child Fam Stud. 1996;5(4):445–67.
79. Dyrborg J, Warborg Larsen F, Nielsen S, Byman J, Buhl Nielsen B, Gautrè-Delay F. The Children's Global Assessment Scale (CGAS) and Global Assessment of Psychosocial Disability (GAPD) in clinical practice - Substance and reliability as judged by intraclass correlations. Eur Child Adolesc Psychiatry. 2000;9(3):195–201.
80. Busner J, Targum SD. The clinical global impressions scale: applying a research tool in clinical practice. Psychiatry (Egmont). 2007;4(7):28–37.
81. Bird HR, Shaffer D, Fisher MS, Gould BS, Chen JY, Hoven C. The Columbia Impairment Scale (CIS): Pilot findings on a measure of global impairment for children and adolescents. Int J Methods Psychiatr Res. 1993;3(3):167–76.
82. Whiteside SP. Adapting the sheehan disability scale to assess child and parent impairment related to childhood anxiety disorders. J Clin Child Adolesc Psychol. 2009;38(5):721–30.
83. Gowers SG, Harrington RC, Whitton A, Lelliott P, Beevor A, Wing J, et al. Brief scale for measuring the outcomes of emotional and behavioural disorders in children. Health of the Nation Outcome Scales for Children and Adolescents (HoNOSCA). Br J Psychiatry. 1999;174(May):413–6.
84. Fabiano GA, Pelham WE, Waschbusch DA, Gnagy EM, Lahey BB, Chronis AM, et al. A practical measure of impairment: Psychometric properties of the impairment rating scale in samples of children with attention deficit hyperactivity disorder and two school-based samples. J Clin Child Adolesc Psychol. 2006;35(3):369–85.
85. Miller SD, Duncan BL, Brown J, Sparks JA, Claud DA. The Outcome Rating Scale: A Preliminary Study of the Reliability, Validity, and Feasibility of a Brief Visual Analog Measure. J Br Ther. 2003;2(2):91–100.
86. Goodman R. The extended version of the Strengths and Difficulties Questionnaire as a guide to child psychiatric caseness and consequent burden. J Child Psychol Psychiatry. 1999 Jul;40(5):791–9.
87. Ermel J, Carter CS, Gold JM, MacDonald AW, Daniel Ragland J, Silverstein SM, et al. Self versus informant reports on the specific levels of functioning scale: Relationships to depression and cognition in schizophrenia and schizoaffective disorder. Schizophr Res Cogn. 2017;9:1–7.
88. Sparrow S, Cicchetti D. The Vineland Adaptive Behavior Scales. 1989;
89. Langley AK, Falk A, Peris T, Wiley JF, Kendall PC, Ginsburg G, et al. The Child Anxiety Impact Scale: Examining Parent- and Child-Reported Impairment in Child Anxiety Disorders. J Clin Child Adolesc Psychol. 2014;43(4):579–91.
90. Lyneham HJ, Sburlati ES, Abbott MJ, Rapee RM, Hudson JL, Tolin DF, et al. Psychometric properties of the Child Anxiety Life Interference Scale (CALIS). J Anxiety Disord. 2013;27(7):711–9.
91. Piacentini J, Peris TS, Bergman RL, Chang S, Jaffer M. Functional impairment in childhood OCD: development and psychometrics properties of the Child Obsessive-Compulsive Impact Scale-Revised (COIS-R). J Clin Child Adolesc Psychol. 2007;36(4):645–53.
92. Ruperto N, Ravelli A, Pistorio A, Malattia C, Cavuto S, Gado-West L, et al. Cross-cultural adaptation and psychometric evaluation of the Childhood Health Assessment Questionnaire (CHAQ) and the Child Health Questionnaire (CHQ) in 32 countries. Review of the general methodology. Clin Exp Rheumatol. 2001;19(4 SUPPL. 23):S1-9.
93. Furber G, Segal L. The validity of the Child Health Utility instrument (CHU9D) as a routine outcome measure for use in child and adolescent mental health services. Health Qual Life Outcomes. 2015;13(1):22.
94. Wille N, Badia X, Bonsel G, Burström K, Cavrini G, Devlin N, et al. Development of the EQ-5D-Y: A child-friendly version of the EQ-5D. Qual Life Res. 2010;19(6):875–86.
95. The Kidscreen Group Europe. The KIDSCREEN questionnaires: Quality of life questionnaires for children and adolescents. Handbook. Lengerich. 2006.
96. Bullinger M, Brütt AL, Erhart M, Ravens-Sieberer U. Psychometric properties of the KINDL-R questionnaire: Results of the BELLA study. Eur Child Adolesc Psychiatry. 2008 Dec;17(SUPPL. 1):125–32.
97. Varni JW, Seid M, Rode CA. The PedsQL™: Measurement model for the pediatric quality of life inventory. Med Care. 1999;37(2):126–39.
98. Endicott J, Nee J, Yang R, Wohlberg C. Pediatric Quality of Life Enjoyment and Satisfaction Questionnaire (PQ-LES-Q): Reliability and validity. J Am Acad Child Adolesc Psychiatry. 2006;45(4):401–7.
99. Forrest CB, Bevans KB, Pratiwadi R, Moon J, Teneralli RE, Minton JM, et al. Development of the PROMIS®

pediatric global health (PGH-7) measure. Qual Life Res. 2014;23(4):1221–31.
100. Brazier JE, Harper R, Jones NMB, O'Cathain A, Thomas KJ, Usherwood T, et al. Validating the SF-36 health survey questionnaire: New outcome measure for primary care. Br Med J. 1992;305(6846):160–4.
101. Bergner M, Bobbitt RA, Carter WB, Gilson BS. The Sickness Impact Profile: Development and Final Revision of a Health Status Measure. Med Care. 1981;19(8):787–805.
102. Scorza P, Stevenson A, Canino G, Mushashi C, Kanyanganzi F, Munyanah M, et al. Validation of the "World Health Organization Disability Assessment Schedule for Children, WHODAS-Child" in Rwanda. PLoS One. 2013;8(3).
103. Group W. Development of the WHOQOL: Rationale and Current Status. Int J Ment Health. 1994;23(3):24–56.
104. Salum GA, Patrick DL, Isolan LR, Manfro GG, De Almeida Fleck MP. Youth quality of life instrument-research version (YQOL-R): Psychometric properties in a community sample. J Pediatr (Rio J). 2012;88(5):443–8.
105. Emslie GJ, Kennard BD, Mayes TL. Predictors of treatment response in adolescent depression. Psychiatr Ann. 2011;41(4):212–9.
106. Knight A, McLellan L, Jones M, Hudson J. Pre-treatment Predictors of Outcome in Childhood Anxiety Disorders: A Systematic Review. Psychopathol Rev. 2014;a1(1):77–129.
107. Lundkvist-Houndoumadi I, Hougaard E, Thastum M. Pre-treatment child and family characteristics as predictors of outcome in cognitive behavioural therapy for youth anxiety disorders. Nord J Psychiatry. 2014;68(8):524–35.
108. Nanni V, Uher R, Danese A. Childhood maltreatment predicts unfavorable course of illness and treatment outcome in depression: A meta-analysis. Am J Psychiatry. 2012;169(2):141–51.
109. Nilsen TS, Eisemann M, Kvernmo S. Predictors and moderators of outcome in child and adolescent anxiety and depression: A systematic review of psychological treatment studies. Eur Child Adolesc Psychiatry. 2013;22(2):69–87.
110. Skriner LC, Chu BC, Kaplan M, Bodden DHM, Bögels SM, Kendall PC, et al. Trajectories and predictors of response in youth anxiety CBT: Integrative data analysis. J Consult Clin Psychol. 2019;87(2):198–211.
111. Walczak M, Ollendick T, Ryan S, Esbjørn BH. Does comorbidity predict poorer treatment outcome in pediatric anxiety disorders? An updated 10-year review. Clin Psychol Rev. 2018;60:45–61.
112. Parkinson S, Eatough V, Holmes J, Stapley E, Midgley N. Framework analysis: a worked example of a study exploring young people's experiences of depression. Qual Res Psychol. 2016;13(2):109–29.
113. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol. 2006;3(2):77–101.
114. Ebesutani C, Reise SP, Chorpita BF, Ale C, Regan J, Young J, et al. The Revised Child Anxiety and Depression Scale-Short Version: Scale reduction via exploratory bifactor modeling of the broad anxiety factor. Psychol Assess. 2012;24(4):833–45.
115. Posner K, Brown GK, Stanley B, Brent DA, Yershova K V., Oquendo MA, et al. The Columbia-suicide severity rating scale: Initial validity and internal consistency findings from three multisite studies with adolescents and adults. Am J Psychiatry. 2011;168(12):1266–77.
116. Ravens-Sieberer U, Erhart M, Rajmil L, Herdman M, Auquier P, Bruil J, et al. Reliability, construct and criterion validity of the KIDSCREEN-10 score: A short measure for children and adolescents' well-being and health-related quality of life. Qual Life Res. 2010;19(10):1487–500.
117. Shaffer D, Gould MS, Brasic J, Fisher P, Aluwahlia S, Bird H. A Children's Global Assessment Scale (CGAS). Arch Gen Psychiatry. 1983;40(11):1228–31.
118. Piqueras JA, Martín-Vivar M, Sandin B, San Luis C, Pineda D. The Revised Child Anxiety and Depression Scale: A systematic review and reliability generalization meta-analysis. J Affect Disord. 2017;218:153–69.
119. Chorpita BF, Moffitt CE, Gray J. Psychometric properties of the Revised Child Anxiety and Depression Scale in a clinical sample. Behav Res Ther. 2005;43(3):309–22.
120. de Ross RL, Gullone E, Chorpita BF. The Revised Child Anxiety and Depression Scale: A Psychometric Investigation with Australian Youth. Behav Chang. 2002;19(2):90–101.
121. Esbjørn BH, Sømhovd MJ, Turnstedt C, Reinholdt-Dunne ML. Assessing the revised child anxiety and depression scale (RCADS) in a national sample of Danish youth aged 8-16 years. PLoS One. 2012;7(5):e37339.
122. Kösters MP, Chinapaw MJM, Zwaanswijk M, van der Wal MF, Koot HM. Structure, reliability, and validity of the revised child anxiety and depression scale (RCADS) in a multi-ethnic urban sample of Dutch children. BMC Psychiatry. 2015 Dec 23;15(1):132.
123. Ebesutani C, Korathu-Larson P, Nakamura BJ, Higa-McMillan C, Chorpita BF. The Revised Child Anxiety

and Depression Scale 25–Parent Version: Scale Development and Validation in a School-Based and Clinical Sample. Assessment. 2017;24(6):712–28.
124. Bagheri Z, Akbari M, Jafari P, Huedo-medina TB. Assessing the measurement invariance of the RCADS-25 questionnaire across gender and child–parent dyads in the presence of multilevel data. Qual Life Res. 2019;28(3):829–40.
125. Klaufus L, Verlinden E, Van Der Wal M, Kösters M, Cuijpers P, Chinapaw M. Psychometric evaluation of two short versions of the Revised Child Anxiety and Depression Scale. BMC Psychiatry. 2020;20(1):1–12.
126. Rodríguez-Jiménez T, Piqueras JA, Lázaro L, Moreno E, Ortiz AG, Godoy A. Metric invariance, reliability, and validity of the Child Version of the Obsessive Compulsive Inventory (OCI-CV) in community and clinical samples. J Obsessive Compuls Relat Disord. 2016;9:1–8.
127. Rosa-Alcázar AI, Ruiz-García B, Iniesta-Sepúlveda M, López-Pina JA, Rosa-Alcázar Á, Parada-Navas JL. Obsessive Compulsive Inventory-Child Version (OCI-CV) in a Spanish community sample of children and adolescents. Psicothema. 2014;26(2):174–9.
128. Martínez-González AE, Rodríguez-Jiménez T, Piqueras JA, Vera-Villarroel P, Godoy A. Psychometric properties of the obsessive-compulsive Inventory-Child Version (OCI-CV) in Chilean children and adolescents. PLoS One. 2015;10(8):1–12.
129. Jones AM, De Nadai AS, Arnold EB, McGuire JF, Lewin AB, Murphy TK, et al. Psychometric properties of the obsessive compulsive inventory: Child version in children and adolescents with obsessive-compulsive disorder. Child Psychiatry Hum Dev. 2013;44(1):137–51.
130. Opakunle T, Aloba O, Akinsulore A. Obsessive compulsive inventory – child version (OCI-CV): Confirmatory factor analysis, reliability, validity and correlates among nigerian adolescents. Malawi Med J. 2018;30(4):262–9.
131. Verlinden E, van Meijel EPM, Opmeer BC, Beer R, de Roos C, Bicanic IAE, et al. Characteristics of the children's revised impact of event scale in a clinically referred dutch sample. J Trauma Stress. 2014;27(3):338–44.
132. Deeba F, Rapee RM, Prvan T. Psychometric properties of the children's revised impact of events scale (CRIES) with Bangladeshi children and adolescents. PeerJ. 2014;2014(1):e536.
133. Salari R, Malekian C, Linck L, Kristiansson R, Sarkadi A. Screening for PTSD symptoms in unaccompanied refugee minors: A test of the CRIES-8 questionnaire in routine care. Scand J Public Health. 2017;45(6):605–11.
134. Gigengack MR, Hein IM, Lindeboom R, Lindauer RJL. Increasing Resource Parents' Sensitivity towards Child Posttraumatic Stress Symptoms: a Descriptive Study on a Trauma-Informed Resource Parent Training. J Child Adolesc Trauma. 2019;12(1):23–9.
135. Ravens-Sieberer U, Gosch A, Rajmil L, Erhart M, Bruil J, Duer W, et al. KIDSCREEN-52 quality-of-life measure for children and adolescents. Expert Rev Pharmacoeconomics Outcomes Res. 2005 Jun;5(3):353–64.
136. Nik-Azin A, Shairi MR, Naeinian MR, Sadeghpour A. The health-related quality of life index kidscreen-10: Confirmatory factor analysis, convergent validity and reliability in a sample of Iranian students. Child Indic Res. 2014;7(2):407–20.
137. Stevanovic D, Tadic I, Novakovic T, Kisic-Tepavcevic D, Ravens-Sieberer U. Evaluating the Serbian version of the KIDSCREEN quality-of-life questionnaires: Reliability, validity, and agreement between children's and parents' ratings. Qual Life Res. 2013;22(7):1729–37.
138. Schorre BEH, Vandvik IH. Global assessment of psychosocial functioning in child and adolescent psychiatry: A review of three unidimensional scales (CGAS, GAF, GAPD). Eur Child Adolesc Psychiatry. 2004;13(5):273–86.
139. Sourander A, Helenius H, Piha J. Child psychiatric short-term inpatient treatment: CGAS as follow-up measure. Child Psychiatry Hum Dev. 1996;27(2):93–104.
140. Orgilés M, Fernández-Martínez I, Morales A, Melero S, Espada JP. Spanish Validation of the Child Anxiety Life Interference Scale (CALIS-C): Psychometric Properties, Factorial Structure and Factorial Invariance Across Gender. Child Psychiatry Hum Dev. 2019;50(5):756–63.
141. Marques T, Pereira AI, Pedro M, Russo V, Goes AR, Barros L. Preliminary analysis of the Portuguese version of the child anxiety life interference scale. Psicol Reflex e Crit. 2015;28(3):454–62.
142. Jones M, Hopkins K, Kyrke-Smith R, Davies R, Vostanis P, Wolpert M. Current view tool: Completion guide. London, United Kingdom: CAMHS Press; 2013.
143. Ebesutani C, Bernstein A, Nakamura BJ, Chorpita BF, Weisz JR. A psychometric analysis of the revised child anxiety and depression scale-parent version in a clinical sample. J Abnorm Child Psychol. 2010;38(2):249–60.