The challenge of ensuring affordability, sustainability, consistency, and adaptability in the common metrics agenda

Karolin Rose Krause, Sophie Chung, Maria da Luz Sousa Fialho, Peter Szatmari, Miranda Wolpert

Mental health research grapples with research waste and stunted field progression caused by inconsistent outcome measurement across studies and clinical settings, which means there is no common language for considering findings. Although recognising that no gold standard measures exist and that all existing measures are flawed in one way or another, anxiety and depression research is spearheading a common metrics movement to harmonise measurement, with several initiatives over the past 5 years recommending the consistent use of specific scales to allow read-across of measurements between studies. For this approach to flourish, however, common metrics must be acceptable and adaptable to a range of contexts and populations, and global access should be as easy and affordable as possible, including in low-income countries. Within a measurement landscape dominated by fixed proprietary measures and with competing views of what should be measured, achieving this goal poses a range of challenges. In this Personal View, we consider tensions between affordability, sustainability, consistency, and adaptability that, if not addressed, risk undermining the common metrics agenda. We outline a three-pronged way forward that involves funders taking more direct responsibility for measure development and dissemination; a move towards managing measure dissemination and adaptation via open-access measure hubs; and transitioning from fixed questionnaires to item banks. We argue that now is the time to start thinking of mental health metrics as 21st century tools to be co-owned and co-created by the mental health community, with support from dedicated infrastructure, coordinating bodies, and funders.

Lancet Psychiatry 2021

Published **Online**
October 14, 2021
https://doi.org/10.1016/S2215-0366(21)00122-X

Cundill Centre for Child and Youth Depression, Centre for Addiction and Mental Health, Toronto, ON, Canada (K R Krause PhD, Prof P Szatmari MD)**; Research Department for Clinical, Educational and Health Psychology, University College London, London, UK** (K R Krause, Prof M Wolpert DClinPsy)**; London, UK** (S Chung BA)**; International Consortium for Health Outcomes Measurement, London, UK** (M d L Sousa Fialho DPhil)**; Department of Psychiatry, University of Toronto, ON, Canada** (Prof P Szatmari)**; Hospital for Sick Children, Toronto, ON, Canada** (Prof P Szatmari)**; Wellcome Trust, London, UK** (Prof M Wolpert)

Correspondence to:
Dr Karolin R Krause, Cundill Centre for Child and Youth Depression, Centre for Addiction and Mental Health, Toronto, ON M6J 1H4, Canada
karolin.krause@camh.ca

Introduction

The design and provision of evidence-based mental health care depends on the availability of reliable, valid, and clinically relevant outcome data—both in research studies and as part of measurement-based care.[1] Although much of medical research relies on established biomarkers or accepted metrics (eg, the body-mass index), reliable biomarkers are yet to be identified for most common mental health conditions.[2] In the absence of such biomarkers, various psychometric scales serve to record symptom clusters, functional impairment, quality of life, or overall wellbeing.

More than 280 scales have been developed over the past century to detect depressive symptoms alone.[3] One scoping review of 30 trials of depression treatments in adolescents identified 19 different instruments used to measure depression symptom severity,[4] and another review identified 30 instruments used to assess anxiety symptoms in children and adolescents across 257 clinical trials and observational studies.[5] A high degree of inconsistency exists between the type of symptoms assessed by different scales.[6,7] A review of 126 questionnaires used to screen for common mental health conditions showed low rates of cross-scale symptom similarity, which ranged from 29% for bipolar disorder to a maximum of 58% for obsessive-compulsive disorder.[7] The psychometric properties of different instruments also vary, as do assessment timelines and the individuals who are asked to report on the phenomenon of interest (ie, clinicians, parents, and children or adolescents). The resulting convoluted evidence base hampers the synthesis and comparability of research findings via meta-analyses or pooled data analysis and the benchmarking of outcomes across services or systems, culminating in research waste and in the stunting of progress in mental health research, including for common conditions such as anxiety and depression.[2,4,8–10]

Several initiatives have aimed to overcome this state of fragmentation by recommending core metrics or core outcome sets that should be administered, as a minimum, across all research studies or practice settings for a given condition.[11,12] The International Consortium for Health Outcomes Measurement has convened several working groups to develop core outcome sets for use in mental health-care settings. As of September, 2021, International Consortium for Health Outcomes Measurement sets were available for anxiety and depression in children, adolescents, and adults,[5,13] as well as for psychotic disorders,[14] personality disorders,[15] and addiction in adolescents and adults.[16] Efforts to develop core outcome sets specifically for clinical trials of treatments for depression in adolescence and adulthood are ongoing.[17,18] In addition, leading mental health funders have agreed on a set of common metrics for mental health that should be measured in all studies done with their support,[19] and UNICEF has led an initiative to identify, adapt, and validate consensus measures for adolescent mental health for use in population surveys worldwide.[20]

Scales recommended by core outcome sets and similar initiatives are typically selected on the basis of specific feasibility criteria, psychometric criteria, or both. The exact criteria can vary depending on the intended use context (eg, clinical trial *vs* measurement-based care), but they often include a consideration of affordability. Affordability is an important factor in ensuring the widest possible uptake of common metrics because cost

	Use incurs a fee*	Controls are placed on tool adaptation, translation, or dissemination	Characteristics of the dissemination model
Beck Youth Inventories (managed by Pearson)[26]	Yes	Yes	Cost and control: a limited set of manuals and questionnaires can be purchased from the copyright holder at a cost; permission must be obtained to reproduce any copyrighted material[27]
HoNOSCA (managed by the Royal College of Psychiatrists)[28]	Case by case	Yes	Partially free and limited use: the free use, copy, and reproduction of HoNOSCA materials without express permission is allowed for care providers within the UK's National Health Service and "in other countries such as Australia, New Zealand and Switzerland, where HoNOS [Health of the Nation Outcome Scales] has been mandated for use to support assessment and outcome monitoring in public and private sector mental health services";[29] otherwise, explicit permission must be obtained from the Royal College of Psychiatrists; developer consent is needed to copy, distribute, or adapt the scale for non-commercial use ("The following acts may not be performed without the consent of the Royal College of Psychiatrists: copying the work; renting or otherwise issuing copies of the work to the public; adapting the work [e.g. changes to the wording of items, scaling of items, addition or deletion of items or changes in the order of items]")[29]
GAD-7 and PHQ-9 (managed by Pfizer)[30,31]	No	No	Free for all users and uses: tool is in the public domain and free for any use ("All PHQ, GAD-7 screeners and translations are downloadable from this website and no permission is required to reproduce, translate, display or distribute them")[32]
WHODAS 2.0 (managed by WHO)[33]	No	Yes	Free use, limited usage: free for non-commercial use, but users are required to sign a user agreement; written permission required for modification, adaptation, or translation ("User shall not modify, abridge, condense, translate, adapt, recast or transform the WHODAS 2.0 in any manner or form, including but not limited to any minor or significant change in wording or organization, or administration procedures, of the WHODAS 2.0. If User thinks that changes are necessary for its work, or if translation is necessary, User must obtain written approval from WHO in advance of making such changes")[34]
PROMIS (managed by a network of primary research sites and coordinating centres)[35]	No	Yes	Free use, collaborative ongoing development: "All English and Spanish versions of PROMIS […] are publicly available for [single] use without licensing or royalty fees for individual research or individual clinical use"; "User agrees not to adapt, alter, amend, abridge, modify, condense, make derivative works, or translate HealthMeasures Instruments without prior written permission from the Provider. In cases where permission is granted, User will be expected to evaluate the impact of approved modifications"; "[…] clinical researchers are encouraged to submit de-identified data for collaborative analysis and reporting. […] Clinical researchers are strongly encouraged to collaborate with HealthMeasures investigators when applying these items and banks to their research"[36]

Data shown as of September, 2021. GAD-7=Generalized Anxiety Disorder 7-item Scale. HoNOSCA=Health of the Nation Outcome Scales for Children and Adolescents. PHQ-9=Patient Health Questionnaire-9. PROMIS=Patient-Reported Outcomes Measurement Information System. WHODAS 2.0=World Health Organization Disability Assessment Schedule 2·0. *Refers to non-commercial use.

Table: **Examples of mental health outcome measurement scales and levels of cost and control**

is one important barrier to the implementation of measurement-based care in practice settings,[21–24] and can also influence the selection of measurement scales in clinical trials or population surveys.

All authors of this Personal View have been involved in one or several of the above-mentioned common metrics initiatives. Through this involvement and conversations with tool developers, we have come to realise that within a measurement landscape dominated by fixed, proprietary, and copyrighted scales, the double aim of promoting consistency and affordability brings inherent tensions related to the adaptability and sustainability of common metrics. We herein discuss how these tensions affect current models of measure development and dissemination, and how they could ultimately undermine the goals of the common metrics agenda. We then suggest three possible ways forward. Although these questions are of relevance to the wider field of mental health research, we will focus on anxiety and depression metrics for children and adolescents as a case example.

A landscape of fixed and proprietary measures

Measurement scales typically consist of a fixed set of items that can be summed to a total score (or several subscale scores) and are usually copyrighted to guarantee developers adequate monetary or non-monetary (eg, attribution) compensation in exchange for allowing others to use their intellectual property.[25] Scales come with varying conditions for reuse, modification, translation, and further dissemination (table). Reliable costing and licencing information is often difficult to find and can be hidden behind the paywalls of commercial measure catalogues. Researchers and practitioners who use or adapt a measure without adhering to licencing terms risk legal persecution, barriers to publishing the resulting research, or calls to retract already published research.[37,38]

Affordability and sustainability

Those seeking to select a measure for use in a research study or measurement-based care system in mental health face the challenge of identifying the most suitable tool from a wide array of choices. Psychometric systematic reviews and meta-analyses provide comparisons of measurement properties that can help with scale selection.[39,40] Feasibility of use in a given target context, which includes the affordability of the scale, is another important selection criterion.[24,41] For example, clinical services or community-based providers might have scarce resources (if any) earmarked for measurement-based care.[24] Within universities, research projects might need free measures due to small funding (eg, for student and trainee projects). For researchers and practitioners in low-income and middle-income countries, the cost of commercially available measures can be prohibitive, especially if the tool is priced with the purchasing power of users in high-income countries as a reference. On the grounds of equity and to build a diverse and inclusive evidence base, core metrics for research and measurement-based care should be as affordable as possible.[37]

Many commonly used mental health measures and gold-standard diagnostic tools can be purchased for a fee from commercial publishers. Within child and adolescent psychiatry, for example, the widely used Beck Youth Inventories, Second Edition[26] measures are held by the educational publisher Pearson. As of September, 2021, a starter kit of Beck Youth Inventories, Second Edition tools including a manual and 25 paper-based inventory booklets could be purchased for US$ 360·10.[42] Time, staff, and financial resources are usually required for the initial conception, design, and validation of a measurement scale,[43] and developers might face additional costs associated with a scale's ongoing management and dissemination. By making a scale commercially available, developers shift a part of these costs onto users. User fees can also compensate for the provision of physical administration kits or further measure development, although how exactly licence fees are used is rarely disclosed.

At the same time, several proprietary scales are provided at no cost to non-commercial users by individual developers, research groups, foundations, or organisations (table), and several systematic reviews and repositories provide helpful catalogues of such freely available measures.[24,41] However, so-called free scales vary in the extent to which they are truly free or open for use, with varying restrictions placed on the user's right to modify, translate, or distribute them. Some tools are available in the public domain and may be used, modified, and distributed without limit and without seeking explicit permission to do so, whereas other tools can only be modified or translated with explicit permission from the developers (table). Although the exact terms of use are not always clearly articulated, these scales do not typically require the payment of a user fee by non-commercial users.

The free availability of some measurement scales might convey the impression that there is no marginal cost involved in controlling and coordinating their use. Yet, as previously mentioned, ongoing support and monitoring is often needed to ensure that a measure is used appropriately (ie, in line with licencing terms), in ways that are methodologically sound, and under minimal risk of harm to respondents. Ongoing inputs might be needed to handle user queries and to support and monitor other researchers' efforts to further validate or adapt the scale. Although initial development is often funded through a research grant, activities related to the ongoing dissemination and oversight of measurement scales might not be explicitly funded.

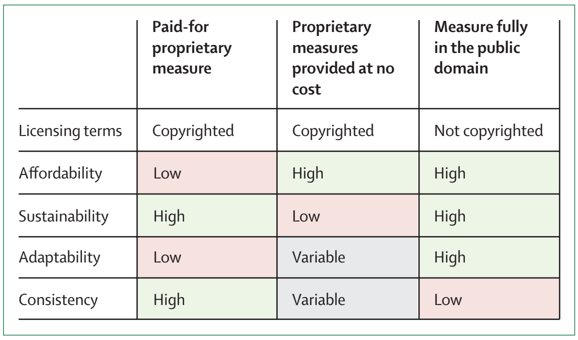
Developers of freely available scales might currently be protected by the fragmentation of the field's attention across the many different existing options. However, if the common metrics movement were to succeed at centring measurement efforts around a small set of free, common metrics, interest and support needs would probably grow exponentially. If no continuous sources of funding are available, a dilemma can ensue for non-commercial providers struggling to meet this increased demand with the infrastructures and resources currently at their disposal. Although the affordability of commercially licenced measures is comparatively poor from the users' point of view, these licencing models are sustainable for providers because they cover ongoing costs, which raises questions about how the field can ensure that a common metrics agenda centred around affordable measures is equitable and sustainable for users and providers alike.

Consistency and adaptability

Common metrics initiatives such as core outcome sets aim to advance the field by integrating and harmonising the evidence base to enable comparisons, benchmarking, and synthesis of data across studies and clinical settings.[5,9,18] Consistent use of a core set of scales (or items) across studies can facilitate the pooling of effect sizes in meta-analyses, and the pooling and linking of datasets as part of integrative data analysis projects that provide enhanced statistical power and allow for the investigation of new research hypotheses.[44–46] In measurement-based care, the tracking of harmonised outcome indicators can enable comparisons between services and mental health systems, thus helping with the identification of best practice examples. To maintain comparability within a landscape of fixed measures, the recommended scales should be used consistently and without substantial modification across studies and settings.

Many mental health scales have been developed in specific clinical settings in high-income countries, and the evidence base relating to their reliability and validity in other contexts (eg, in low-income and middle-income countries, or in population-based surveys) is only gradually emerging. In turn, many measures that assess health-related quality of life were originally developed for use with populations in physical health-care settings, or in non-clinical populations, and might require adaptation for meaningful use in mental health contexts.

More generally, important opportunities exist to strengthen measurement scales on the basis of feedback from researchers, practitioners, and those with lived experience. Currently, measurement is based on imperfect scales, some of which have been widely used for



	Paid-for proprietary measure	Proprietary measures provided at no cost	Measure fully in the public domain
Licensing terms	Copyrighted	Copyrighted	Not copyrighted
Affordability	Low	High	High
Sustainability	High	Low	High
Adaptability	Low	Variable	High
Consistency	High	Variable	Low

***Figure 1:* Trade-offs between measure development and dissemination models in the current status quo**

decades. For example, the Children's Depression Rating Scale–Revised[47] is the symptom measure that is most widely used across trials for treatments of depression in adolescence.[48] However, a systematic review and evidence appraisal done using the COnsensus-based Standards for the selection of health Measurement INstruments guidelines suggests that the evidence supporting the Children's Depression Rating Scale–Revised's measurement properties in adolescents aged 12–18 years with major depression is scarce: only six relevant psychometric studies were identified, and none of these studies assessed content validity, cross-cultural validity and measurement invariance, or measurement error.[49] In many cases, common metrics initiatives are basing their recommendations on measures that are considered good enough, rather than on a gold standard.[2,5] Using existing measurement scales that are considered good enough is a compromise made to kick-start a common metrics movement and accelerate learning about optimal approaches to harmonisation. This compromise is based on the understanding that common metrics should be assessed as a minimum, but can be complemented flexibly with additional scales, which can help to identify any issues, idiosyncrasies, or limitations of the recommended tools.[50] To avoid a premature centring of measurement around tools that might later be shown to have important flaws, piloting them in a variety of populations and contexts and adapting or exchanging them if needed is vital. For example, the implementation of International Consortium for Health Outcomes Measurement outcome sets is overseen by steering groups with a mandate to periodically review evidence from pilot studies and consider whether any measurement recommendations should be revised. At present, however, the fixed nature of most scales is a barrier to adaptation because their revision or replacement can compromise the comparability of data collected across different timepoints or populations.

The proprietary nature of many existing scales is another barrier to adaptability. Commercially licenced measures often cannot be easily modified by users for applications in new populations or contexts. Control over the measure is often firmly centralised with the licence holder, which can mean foregoing opportunities for more dynamic and collective improvement efforts. Although the scope for adaptation is comparatively low, this model provides maximum consistency, ensuring that only official versions of a scale are in circulation. Developers providing their tool at no cost often allow greater leeway for adaptation by considering collaborations, while striving to preserve the highest possible consistency. As previously discussed, this double effort can require considerable resources, especially if the relevant measurement scales find themselves at the heart of a common metrics movement. A third model, in which measures are fully in the public domain and open to modification and adaptation without control from a developer or provider, meets criteria of affordability, sustainability, and adaptability, but provides no assurance of consistent use, which might undermine their utility as common metrics.

An obvious tension exists between the desire to enable iterative adaptation and the need to preserve consistency in a scale's characteristics and administration. Similarly, tension exists between desires for adaptation and the licencing and copyright terms attached to proprietary measures, which might prohibit modifications. These tensions are difficult to reconcile within a framework of fixed and proprietary measures. Figure 1 visually illustrates the authors' subjective understanding of the trade-offs involved in current models of measure development and dissemination in relation to affordability, sustainability, adaptability, and consistency. A low or high rating indicates a comparatively small or large perceived chance of providing the relevant attribute within the given development and dissemination model, whereas a variable rating indicates that whether or not an attribute is likely to be achieved depends on the specificities of the given development and dissemination model.

Where to go from here? A three-pronged way forward

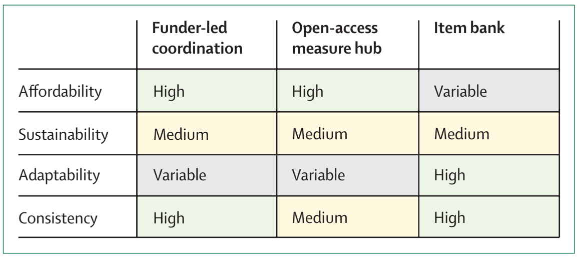
Our view is that the inherent tensions between affordability and sustainability, consistency, and adaptability of fixed proprietary measures pose an important barrier to the successful implementation of the common metrics agenda in mental health. A focus on affordability without consideration of sustainability will probably be ineffective because providers can find themselves unable to uphold free distribution models in the face of increasing demand for user support. In turn, a focus on sustainability without consideration of affordability would fall short on equity of access and the representativeness of the ensuing evidence base.[37] Because the common metrics agenda aims to generate a more inclusive, integrated, and higher quality evidence base while still promoting harmonisation, the competing needs for adaptation and consistency must also be considered. Hereafter, we outline a three-pronged approach as one possible avenue towards providing a better balance of affordability, sustainability, consistency, and adaptability in measure development and dissemination (figure 2), and towards moving the development and dissemination of common metrics into the 21st century.

Changing the funding and dissemination model for proprietary measures

Several funders (eg, National Institutes of Health and Wellcome Trust) require that the funded research findings—and, in some cases, the underlying data—be shared via open access publishing or deposition in accessible databases or repositories.[37] Funders might want to extend such requirements to any measurement tools developed with their support, by either allowing researchers to budget for anticipated dissemination or maintenance costs in their initial grant applications or by developing a mechanism for periodically renewing funding for the maintenance and oversight of freely distributed metrics. Funders might even establish systems whereby individual developers can transfer the responsibility for measure dissemination to the funder, who can then earmark resources and develop mechanisms for ongoing oversight. This model offers a limited scope for dynamic scale adaptation, and is not inherently sustainable because it requires ongoing investment and commitment from funders. It does, however, provide relatively high affordability and scale consistency, and might enhance sustainability in the short term (within a landscape of fixed proprietary scales).

Learning from data science: a measure hub model

We propose that in the longer term, the field might need to shift its mindset from seeing measurement scales as fixed and proprietary, to thinking of measures as evolving tools or code that should be co-owned and co-created by the user community, including researchers, practitioners, and those with lived experience. Opportunities might exist for learning from pioneering work in computer sciences and statistics and from open-science approaches, wherein openly available tools are believed to facilitate high-quality research practices and increase research efficiency.[51] For example, in computer sciences, the Open Science Grid is a consortium of stakeholder communities (eg, researchers, information technology providers, software developers, and educators) that share computing resources to advance scientific practice.[52] Similarly, computer scientists and statisticians freely share code through platforms such as GitHub. Developers can licence their code with various licencing models that provide different levels of control and attribution, including permissive free software licences that impose minimal restrictions on use and distribution (eg, Berkeley Software Distribution licences). The hub enables developers to monitor who is using their code and how it is being adapted. As such, it provides a sophisticated and highly regulated model of version control, which might help to reconcile adaptation and consistency—at least to a degree.



	Funder-led coordination	Open-access measure hub	Item bank
Affordability	High	High	Variable
Sustainability	Medium	Medium	Medium
Adaptability	Variable	Variable	High
Consistency	High	Medium	High

Figure 2: **Trade-offs between different alternative measure development and dissemination models**

An open-access measure hub in mental health could facilitate the sharing of non-commercial measurement scales and their flexible adaptation for use in new settings and populations. The Wikiversity Evidence Based Assessment Portfolios already compile information related to mental health scales, which can be edited and expanded collaboratively by Wikiversity users.[53] However, the portfolios do not currently host the scales themselves, or help with version control or the coordination of adaptation efforts. Other instrument repositories helpfully compile copies of free instruments where available, in addition to providing information on their measurement properties.[24] However, existing repositories might not always be exhaustive or up to date, as highlighted in a recent review: "the measures contained in each repository often did not overlap and were not always updated with the latest versions of the measure. Repositories varied in how they selected measures to include; some required authors to self-submit and self-report on the measure's psychometric evidence, others gathered experts to recommend measures for inclusion. The dynamic nature of these repositories suggests that the landscape of freely available measures may shift quickly; however, in the absence of a single, coordinated effort to house pragmatic measures, these repositories are unlikely to keep pace with advancing science."[24]

To "keep pace with advancing science", a more centralised and dynamic measure hub that can be updated by the user community itself might be needed. A centralised hub model can aid with sustainability by reducing the cost and effort required for overseeing measure dissemination, where the hub provides a transparent track record of use, validation, and adaptation efforts. The onus for providing technical support and answering user queries could be shifted, at least partly, from individual developers to the wider user community through open discussion boards and forums. Finally, a hub model might reduce the risk of inadequate or unauthorised use, by providing clear and transparent licencing terms. On the downside, although enabling more transparent version control, the hub model prioritises adaptation over consistency, meaning that the

For **GitHub** see https://github.com/

number of different versions in circulation would probably increase and that comparability could decline. Furthermore, a developer would not receive monetary compensation for initial development efforts and intellectual property, which might deter some developers from disseminating their tools in this way.

Moving from fixed measures to an item bank model

Although a centralised, open-access measure hub could make the adaptation and use of free measures easier and more dynamic, initial scale creation still lies with individual developers or research teams. A scale might be adaptable, but it would remain a largely fixed tool. An alternative model that maximises the potential for dynamic measure creation is a move from fixed scales to item banks and personalised assessments based on item response theory and computerised adaptive testing.[54]

Item banks contain numerous questionnaire items that, through use of item response theory models, have been calibrated to assess a specific construct, such as depression or anxiety, at a defined level of difficulty (or severity). With computerised adaptive testing, individual items can be flexibly selected to create tailored assessments that locate individuals more precisely and rapidly on the construct continuum of interest than fixed scales that were created using classical test theory.[55,56] Item bank approaches based on computerised adaptive testing enable consistent scoring across varying item combinations. As a result, developing short forms that are tailored to specific populations and still generate scores that are directly comparable becomes possible,[55] thus providing a way out of the adaptability versus consistency dilemma.

For details on the **PROsetta Stone** see https://www.prosettastone.org/

The National Institutes of Health's Patient-Reported Outcomes Measurement Information System (PROMIS) is an item bank that was calibrated to capture patient-reported outcomes across a range of chronic health conditions. PROMIS items are assembled from existing measures via literature searches, reviewed through consultation with key stakeholders (including those with lived experience), and subject to psychometric testing and validation. The PROMIS network is organised around several primary research sites, with a statistical coordinating centre managing the development and validation of items and providing a data management and storage system.

To date, PROMIS has not replaced proprietary fixed measurement scales in much of child and adolescent anxiety and depression research, although relevant item banks and short forms are available.[57] PROMIS measures were developed and calibrated as non-disease-specific scales to assess emotional health (including anxiety and depression) in individuals with chronic health conditions, rather than specifically for the purpose of clinical mental health assessment.[58] In addition, although printout short forms are available, personalised assessment via computerised adaptive testing requires digital administration, which is not (yet) feasible in all contexts. Despite being more sustainable than free measure distribution via individual developers, PROMIS does not have inbuilt sustainability because substantial inputs and resources by the PROMIS network and continued external funding are required to maintain the item bank. There are limitations to affordability compared with freely distributed measures: although English and Spanish language fixed scales were available from PROMIS at no cost as of September, 2021, other translations did incur a distribution fee.[59] A review fee was charged for quality assurance of new translations and for ensuring that they were harmonised across languages. The relatively slow uptake of the PROMIS measures for the purpose of clinical mental health assessment suggests a need to further examine the widespread feasibility of centring common metrics initiatives around item bank and item response theory models.

In the meantime, item response theory offers opportunities to harmonise and compare scores obtained from fixed legacy measures, which could help to smooth the transition from a fixed to a more dynamic measurement landscape. With item response theory, items from existing fixed scales can be calibrated within the same measurement model used by an item bank, which allows linking of the scores of legacy scales to a common underlying scoring metric. The scores of different scales can then be compared via crosswalk tables,[60] or online resources such as PROsetta Stone. Relevant studies have been done using PROMIS as a common metric for several common depression and anxiety scales as part of the PROsetta Stone project.[61–64] Once a linkage is established, comparability can be preserved when moving from one fixed scale to another. The continued development of such models and the linkage of additional scales might eventually offer a perspective for creating common metrics that are independent from individual instruments, without requiring the use of a single common instrument, and without sacrificing the comparability of legacy data. Funders could support this transition by requiring that any newly developed or adapted fixed scales be calibrated onto an existing measurement model.

Looking ahead

To overcome the current state of data fragmentation in mental health research and to promote the creation of a consistent, diverse, and inclusive evidence base while avoiding the emergence of a new line of fragmentation driven by cost, centring research efforts around affordable common metrics is important. Cost-driven fragmentation would also undermine efforts to enhance the quality of the evidence base by making the most suitable tools the purview of a comparatively small group of researchers and practitioners with the necessary financial resources.

Common metrics and Core Outcome Set initiatives in anxiety and depression research[5,13,19,20] have recommended measures such as the Generalized Anxiety Disorder 7-item Scale,[30] the Patient Health Questionnaire-9,[31] the WHO Disability Assessment Schedule 2.0 12-item short form,[33] and the Revised Children's Anxiety and Depression Scale 25-item short version.[65] Although these were selected for being freely available at the time of selection (among other criteria), there is no guarantee that they will be provided at no cost indefinitely, especially if developers or providers begin to feel overburdened by increasing demand for support. Although these measures were considered good enough[2] for the time being, they might display important limitations once piloted across a greater variety of contexts and settings. To base the common metrics movement on the best possible measurement scales, adapting, modifying, and tailoring to specific populations is essential, without undermining the overarching aim of harmonisation. The future of the common metrics agenda might lie in moving beyond fixed proprietary measurement scales towards models that provide greater scope for dynamic adaptation and tailoring, while maintaining the necessary degree of consistency. Item banks, item response theory, and computerised adaptive testing will probably play a central role in this effort.

Now is the time for the the mental health research community to rethink how measure development and dissemination can be organized within a common metrics framework. No single model is likely to resolve the outlined tensions between affordability, sustainability, adaptability, and consistency today. We have discussed a three-pronged way forward that involves revisiting current models for the dissemination of fixed proprietary measures, considering opportunities for the development of more dynamic assessments. Rather than thinking of measures as 20th century manuscripts or instruments, starting to think of them as the equivalent of 21st century tools or computation code that can be co-owned and co-created by the wider research community with support from dedicated infrastructure, coordinating bodies, or funders might be useful. Funders have already started to agree on the key building blocks of a common metrics toolbox.[66] The next step is to help to support the wider mental health science community to start building from these blocks to create free and flexible metrics in ways that are sustainable and also preserve comparability and consistency.

Contributors

KRK, SC, MdLSF, and MW conceived this Personal View. KRK, MdLSF, and MW conducted relevant background research. KRK drafted the manuscript with technical and intellectual input from all authors. All authors contributed to refining the manuscript. All authors read and approved the final version of the manuscript.

Declaration of interests

KRK reports personal fees from the International Consortium for Health Outcomes Measurement (ICHOM), outside the submitted work; was a research fellow with ICHOM and supported the development of a core outcome set for children and adolescents experiencing anxiety and depression between Oct 1, 2018, and March 31, 2020; is involved with the International Network for Research Outcomes in Adolescent Depression Studies initiative at the Hospital for Sick Children (Toronto, ON, Canada) that aims to develop a core outcome set specifically for clinical trials of treatmetns for depression in adolescence; and has been serving as an advisor on the Common Measures in Mental Health Science Advisory Committee convened by the International Alliance of Mental Health Research Funders since April, 2021. SC was an employee of ICHOM between June 25, 2018, and Dec 30, 2019, and was responsible for managing the development of core outcome sets in mental health. MdLSF is employed by ICHOM and oversees the development of core outcome sets in mental health. PS reports royalties from Guilford Press for his book *A Mind Apart: Understanding Children with Autism and Asperger Syndrome*; royalties from Simon & Schuster for his book *Start Here: A Parent's Guide to Helping Children and Teens Through Mental Health Challenges*; and is a co-investigator for the International Network for Research Outcomes in Adolescent Depression Studies initiative that aims to develop a core outcome set specifically for clinical trials of depression in adolescence. MW is Director of Mental Health at the Wellcome Trust and as such has issued recommendations for common metrics in anxiety and depression research. MW chaired (but not voted at) the aforementioned ICHOM working group that developed core metrics for anxiety and depression in children and young people; was formerly Director of the Child Outcomes Research Consortium that made recommendations about outcome measures; and was involved in developing one of the measures recommended by the ICHOM working group (the Current View tool, which is free to use, and results in no financial gain for the author). The authors did not receive any grant or funding that was specifically dedicated to the preparation of this manuscript.

Acknowledgments

KRK is funded by the Cundill Centre for Child and Youth Depression at the Centre for Addiction and Mental Health, in Toronto (ON, Canada). PS has received grant funding support from the Canadian Institutes of Health Research and the Jamie and Patsy Anderson Chair in Child and Youth Mental Health. We thank the reviewers of this manuscript for their valuable feedback and suggestions.

References

1. Harding KJK, Rush AJ, Arbuckle M, Trivedi MH, Pincus HA. Measurement-based care in psychiatric practice: a policy framework for implementation. *J Clin Psychiatry* 2011; **72:** 1136–43.
2. *The Lancet Psychiatry*. A good enough measure. *Lancet Psychiatry* 2020; **7:** 825.
3. Santor DA, Gregus M, Welch A. Focus article: eight decades of measurement in depression. *Measurement* 2006; **4:** 135–55.
4. Mew EJ, Monsour A, Saeed L, et al. Systematic scoping review identifies heterogeneity in outcomes measured in adolescent depression clinical trials. *J Clin Epidemiol* 2020; **126:** 71–79.
5. Krause KR, Chung S, Adewuya AO, et al. International consensus on a standard set of outcome measures for child and youth anxiety, depression, obsessive-compulsive disorder, and post-traumatic stress disorder. *Lancet Psychiatry* 2021; **8:** 76–86.
6. Fried EI. The 52 symptoms of major depression: lack of content overlap among seven common depression scales. *J Affect Disord* 2017; **208:** 191–97.
7. Newson JJ, Hunter D, Thiagarajan TC. The heterogeneity of mental health assessment. *Front Psychiatry* 2020; **11:** 76.
8. 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. *J Am Acad Child Adolesc Psychiatry* 2020; **59:** 810–41.
9. Szatmari P, Offringa M, Butcher NJ, Monga S. Counting what counts: the case for harmonized outcomes in child and youth mental health research. *J Am Acad Child Adolesc Psychiatry* 2019; **58:** 656–58.
10. Monga S, Offringa M, Butcher NJ, Szatmari P. From research to practice: the importance of appropriate outcome selection, measurement, and reporting in pediatric mental health research. *J Am Acad Child Adolesc Psychiatry* 2020; **59:** 497–500.
11. Clarke M, Williamson PR. Core outcome sets and systematic reviews. *Syst Rev* 2016; **5:** 11.

12 Williamson PR, Altman DG, Bagley H, et al. The COMET handbook: version 1.0. *Trials* 2017; **18** (suppl 3)**:** 280.

13 Obbarius A, van Maasakkers L, Baer L, et al. Standardization of health outcomes assessment for depression and anxiety: recommendations from the ICHOM Depression and Anxiety Working Group. *Qual Life Res* 2017; **26:** 3211–25.

14 McKenzie E, Matkin L, Sousa Fialho L, et al. Developing an international standard set of patient-reported outcome measures for psychotic disorders. *Psychiatr Serv* 2021; published online Aug 9. https://doi.org/10.1176/appi.ps.202000888.

15 Prevolnik Rupel V, Jagger B, Fialho LS, et al. Standard set of patient-reported outcomes for personality disorder. *Qual Life Res* 2021; published online June 2. https://doi.org/10.1007/s11136-021-02870-w.

16 ICHOM. Standard sets—mental health. 2021. https://www.ichom.org/standard-sets/#mental-health (accessed Sept 15, 2021).

17 Chevance A, Ravaud P, Tomlinson A, et al. Identifying outcomes for depression that matter to patients, informal caregivers, and health-care professionals: qualitative content analysis of a large international online survey. *Lancet Psychiatry* 2020; **7:** 692–702.

18 Monga S, Monsour A, Stallwood E, et al. Core outcome set development for adolescent major depressive disorder clinical trials: a registered report. *J Am Acad Child Adolesc Psychiatry* 2020; **59:** 1297–98.

19 Wolpert M. Funders agree first common metrics for mental health science. July 6, 2020. https://www.linkedin.com/pulse/funders-agree-first-common-metrics-mental-health-science-wolpert/ (accessed Sept 15, 2021).

20 UNICEF. Measurement of mental health among adolescents at the population level (MMAP): an overview. 2019. https://data.unicef.org/wp-content/uploads/2018/02/Mental-health-measurement_MMAP_overview_UNICEF.pdf (accessed Sept 15, 2021).

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

22 Kotte A, Hill KA, Mah AC, et al. Facilitators and barriers of implementing a measurement feedback system in public youth mental health. *Adm Policy Ment Health* 2016; **43:** 861–78.

23 Whiteside SPH, Sattler AF, Hathaway J, Douglas KV. Use of evidence-based assessment for childhood anxiety disorders in community practice. *J Anxiety Disord* 2016; **39:** 65–70.

24 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:** 1–17.

25 Leval PN. Campbell as fair use blueprint? *Wash Law Rev* 2015; **90:** 597.

26 Beck JS, Beck AT, Jolly J, Steer RA. Beck Youth Inventories—second edition for children and adolescents (BYI-II). San Antonio, TX: Pearson; 2005.

27 Pearson. Clinical permissions & licensing requests. 2021. https://www.pearsonassessments.com/footer/permissions---licensing.html (accessed Sept 15, 2021).

28 Garralda ME, Yates P, Higginson I. Child and adolescent mental health service use. HoNOSCA as an outcome measure. *Br J Psychiatry* 2000; **177:** 52–58.

29 Royal College of Psychiatrists. Health of the Nation Outcome Scales (HoNOS). https://www.rcpsych.ac.uk/events/in-house-training/health-of-nation-outcome-scales (accessed Sept 15, 2021).

30 Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med* 2006; **166:** 1092–97.

31 Spitzer RL, Kroenke K, Williams JBW. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. *JAMA* 1999; **282:** 1737–44.

32 Pfizer. Screener overview. https://www.phqscreeners.com/select-screener (accessed Sept 15, 2021).

33 Ustün TB, Chatterji S, Kostanjsek N, et al. Developing the World Health Organization disability assessment schedule 2.0. *Bull World Health Organ* 2010; **88:** 815–23.

34 WHO. User agreement for WHODAS 2.0. https://www.who.int/classifications/icf/WHODAS_2_0_UserAgreement_v2011.pdf?ua=1 (accessed Sept 15, 2021).

35 Cella D, Riley W, Stone A, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. *J Clin Epidemiol* 2010; **63:** 1179–94.

36 HealthMeasures. HealthMeasures terms of use. December, 2017. http://www.healthmeasures.net/images/PROMIS/Terms_of_Use_HM_approved_1-12-17_-_Updated_Copyright_Notices.pdf (accessed Sept 15, 2021).

37 Hays RD, Weech-Maldonado R, Teresi JA, Wallace SP, Stewart AL. Commentary: copyright restrictions versus open access to survey instruments. *Med Care* 2018; **56:** 107–10.

38 Marcus A. Psychology paper retracted after creators of tool allege “serious breach of copyright”. Dec 29, 2020. https://retractionwatch.com/2020/12/29/psychology-paper-retracted-after-creators-of-tool-allege-serious-breach-of-copyright/ (accessed Sept 15, 2021).

39 Youngstrom EA, Egerton GA, Genzlinger J, Freeman LK, Rizvi SH, Van Meter A. Improving the global identification of bipolar spectrum disorders: meta-analysis of the diagnostic accuracy of checklists. *Psychol Bull* 2018; **144:** 315–42.

40 Stockings E, Degenhardt L, Lee YY, et al. Symptom screening scales for detecting major depressive disorder in children and adolescents: a systematic review and meta-analysis of reliability, validity and diagnostic utility. *J Affect Disord* 2015; **174:** 447–63.

41 Beidas RS, Stewart RE, Walsh L, et al. Free, brief, and validated: standardized instruments for low-resource mental health settings. *Cognit Behav Pract* 2015; **22:** 5–19.

42 Pearson Assessments. Beck Youth Inventories—Second edition overview. https://www.pearsonassessments.com/store/usassessments/en/Store/Professional-Assessments/Personality-%26-Biopsychosocial/Beck-Youth-Inventories-%7C-Second-Edition/p/100000153.html# (accessed Sept 15, 2021).

43 Mokkink LB, Prinsen CA, Patrick D, et al. COSMIN study design checklist for patient-reported outcome measurement instruments. July, 2019. https://gut.bmj.com/content/gutjnl/70/1/139/DC1/embed/inline-supplementary-material-1.pdf?download=true (accessed Sept 15, 2021).

44 Hussong AM, Curran PJ, Bauer DJ. Integrative data analysis in clinical psychology research. *Annu Rev Clin Psychol* 2013; **9:** 61–89.

45 Hofer SM, Piccinin AM. Integrative data analysis through coordination of measurement and analysis protocol across independent longitudinal studies. *Psychol Methods* 2009; **14:** 150–64.

46 Bainter SA, Curran PJ. Advantages of integrative data analysis for developmental research. *J Cogn Dev* 2015; **16:** 1–10.

47 Poznanski E, Mokros H. Children’s Depression Rating Scale–Revised (CDRS-R). Los Angeles, CA: Western Psychological Services, 1996.

48 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:** 61–71.

49 Stallwood E, Monsour A, Rodrigues C, et al. Systematic review: the measurement properties of the Children’s Depression Rating Scale-Revised in adolescents with major depressive disorder. *J Am Acad Child Adolesc Psychiatry* 2021; **60:** 119–33.

50 Patalay P, Fried EI. Editorial perspective: prescribing measures: unintended negative consequences of mandating standardized mental health measurement. *J Child Psychol Psychiatry* 2021; **62:** 1032–36.

51 Bartling S, Friesike S. Opening Science. Cham: Springer International Publishing, 2014.

52 Altunay M, Avery P, Blackburn K, et al. A science driven production cyberinfrastructure—the Open Science Grid. *J Grid Comput* 2011; **9:** 201–18.

53 No authors listed. Evidence-based assessment/Assessment portfolios V2. 2021. https://en.wikiversity.org/wiki/Evidence-based_assessment/Assessment_portfolios_V2#Instructions_for_Editing_a_Portfolio (accessed Sept 15, 2021).

54 Thissen D, Steinberg L. Item response theory. In: Millsap RE, Maydeu-Olivares A, eds. The Sage handbook of quantitative methods in psychology. London: Sage Publications; 2009: 148–77.

55 Amtmann D, Cook KF, Jensen MP, et al. Development of a PROMIS item bank to measure pain interference. *Pain* 2010; **150:** 173–82.

56 Revicki DA, Cella DF. Health status assessment for the twenty-first century: item response theory, item banking and computer adaptive testing. *Qual Life Res* 1997; **6:** 595–600.

57 Irwin DE, Stucky B, Langer MM, et al. An item response analysis of the pediatric PROMIS anxiety and depressive symptoms scales. *Qual Life Res* 2010; **19:** 595–607.

58 DeWalt DA, Gross HE, Gipson DS, et al. PROMIS pediatric self-report scales distinguish subgroups of children within and across six common pediatric chronic health conditions. *Qual Life Res* 2015; **24:** 2195–208.

59 HealthMeasures. PROMIS: obtain and administer measures. 2021. https://www.healthmeasures.net/explore-measurement-systems/promis/obtain-administer-measures (accessed Sept 15, 2021).

60 Fischer HF, Rose M. www.common-metrics.org: a web application to estimate scores from different patient-reported outcome measures on a common scale. *BMC Med Res Methodol* 2016; **16:** 142.

61 Choi SW, Schalet B, Cook KF, Cella D. Establishing a common metric for depressive symptoms: linking the BDI-II, CES-D, and PHQ-9 to PROMIS depression. *Psychol Assess* 2014; **26:** 513–27.

62 Kaat AJ, Newcomb ME, Ryan DT, Mustanski B. Expanding a common metric for depression reporting: linking two scales to PROMIS depression. *Qual Life Res* 2017; **26:** 1119–28.

63 Kaat AJ, Kallen MA, Nowinski CJ, Sterling SA, Westbrook SR, Peters JT. PROMIS pediatric depressive symptoms as a harmonized score metric. *J Pediatr Psychol* 2020; **45:** 271–80.

64 Schalet BD, Cook KF, Choi SW, Cella D. Establishing a common metric for self-reported anxiety: linking the MASQ, PANAS, and GAD-7 to PROMIS Anxiety. *J Anxiety Disord* 2014; **28:** 88–96.

65 Ebesutani C, Reise SP, Chorpita BF, 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:** 833–45.

66 International Alliance of Mental Health Research Funders. 2020 year in review. 2020. https://iamhrf.org/sites/iamhrf.org/files/uploads/page/files/year_in_review_2021.pdf (accessed Sept 15, 2021).