Abstract

Objective: Measurement of treatment outcomes in childhood depression has traditionally focused on assessing symptoms from the clinician's perspective, without exploring other outcome domains or considering young people's perspectives. This systematic review explored the extent to which multidimensional and multi-informant outcome measurement have been used in clinical research for adolescent depression in the past decade, and how patterns evolved over time.

Method**:** We searched Embase, MEDLINE and PsycINFO, and included studies published between 2007 and 2017 that assessed the effectiveness of treatments or service provision for adolescent depression. Of 7483 studies screened, 95 met inclusion criteria and were included for data extraction and analysis.

Results**:** We identified ten outcomes domains, of which studies assessed two on average, using four outcome measures. Most studies (94%) measured symptoms, followed by functioning (52%). Other domains such as personal growth, relationships, quality of life, or service satisfaction were each considered by less than one in ten studies. Youth self-report was considered by 54% but tended to focus on secondary outcomes. Multidimensional and multi-informant measurement has been more frequent in studies published since 2015.

Conclusion: Recent clinical research continues to prioritize symptoms measurement based on clinician report and has not yet fully embraced multi-dimensional and multi-informant approaches. In the context of significant policy shifts towards patient-centered and evidence-based care, measuring what matters most to patients has become a priority, but this is not yet widely reflected in clinical research.

Key Words: depression, adolescent, outcome measure, taxonomy, outcome assessment.

Introduction

Over the past two decades, health care systems around the world have come under pressure due to demographic change and spiralling costs.[1] Governments responded by accelerating efforts to provide high quality care at the lowest possible cost, through evidence-based practice, quality improvement, and value-based purchasing.[2,3] These shifts have placed the measurement of health outcomes at the heart of political decision-making, planning and commissioning. The UK Department of Health has described the session-by-session monitoring of treatment outcomes as “one of the most important innovations in psychological therapies”,[4(p2)] and the governments of the United States (US) and Australia have introduced core sets of outcome measures for children to promote consistent outcome measurement across states and territories.[5,6]

The US core set includes a measure relating to adolescent depression amongst 26 core measures of child health, which reflects the significant burden that depression represents in the US and around the wold. Indeed, the World Health Organization describes unipolar depression as the greatest contributor to the burden of disease in young people aged 12-19 worldwide.[7] The US National Committee for Quality Assurance (NCQA) will launch a Depression Quality Improvement Collaborative in the summer of 2018, specifically to promote quality and outcomes assessment in the treatment of adolescent depression.[8]

Never has there been a similar level of attention to what, how, and whether mental health outcomes are changing, and never has outcome measurement played such an important role in the allocation of funds and design of services. In this context, the question of what constitutes a “good” outcome is a pressing priority. Yet, there is currently no consensus on what outcomes are most important, to whom, and in what context.

Traditionally, policy and practice have focused on symptom change, generally determined by clinical judgement, as the locus of attention.[9–11] However, understandings of what makes for a ‘good outcome’ are far from clear-cut and shaped by factors such as political economy and ideology,[12] scientific paradigms,[13] and societal views of what constitutes a valuable citizen.[14–16] Young people and their parents commonly do not share views about treatment goals, with children prioritizing internalizing difficulties and personal growth, and parents focusing on externalizing problems and strategies for behavior management.[17,18] In order to clarify what constitutes a ‘good outcome’ in clinical research and practice in relation to adolescent depression, two key issues must be considered: the issue of outcome domains of focus and the issue of multiple perspectives.

In terms of domains of focus, it is not clear how far a positive outcome should be defined exclusively in terms of symptom change or with respect to a broader range of outcome domains. Consultations with service users of all ages have highlighted the value they accord to measuring change beyond symptoms, in domains such as interpersonal relationships, ability to cope, and having a sense of control over one’s life.[19–21] A thematic analysis of young peoples’ self-defined goals recorded by English child mental health services identified three overarching domains relevant to young service users: *relationships; coping with problems and symptoms; and personal growth and functioning.*[17] A conceptual model based on theoretical considerations about the dynamic nature of childhood and the importance of environmental factors suggests that outcomes should be measured across seven domains, namely *symptoms, functioning, consumer perspectives, environments systems, parental symptoms and physical health.*[10,22]

Turning to the second key challenge, that of difference in perspective: differences in perceptions of change between reporters are common in relation to childhood depression, as highlighted by a recent meta-analysis which stressed that “different informants observe different samples of a youth’s behavior, in different contexts, and bring different perspectives to what they observe”.[23] There is an increased call for greater youth participation in child mental health treatment including determining which outcomes to assess,[4,24–26] and a growing awareness that high integrity health care systems cannot simply rely on clinicians' views of what is most important, or on symptom change alone, but must consider patients' perspectives to ensure that treatment meets their needs, priorities, and preferences.[27]

The current study

The present study aims to establish the extent to which multidimensional and multi-informant approaches to outcome measurements have been used in clinical research for adolescent depression over the past ten years and whether patterns are changing. It builds upon previous reviews, which were broad in terms of covering a wide range of disorders and developmental stages, and specific in considering only controlled studies.[9,10,28,29] This review focuses on one specific disorder and developmental stage (adolescent depression), while broadening the scope to consider both controlled and non-controlled studies. While previous reviews identified what overarching outcome domains were covered, this review provides additional insight by summarizing the sub-concepts measured within each domain, the measures used, and the extent to which they consulted young people. The four research questions guiding this review are: 1) What outcome domains and subdomains have been measured for treatments of adolescent depression between 2007 and 2017? 2) What outcome measures were used? 3) To what extent where the views of young people considered? 4) To what extent have patterns changed over time?

Method

Search strategy

Three academic databases, PsycINFO, Medline and Embase, were systematically searched for studies reporting outcomes of depression treatments for adolescents. Studies had to be published in peer-reviewed journals between January 2007 and July 2017. We used Boolean operators to identify pre-specified search terms in titles and abstracts that described the study population (e.g. “child”, “adolescent*”), the disorder (“mental health”, “depress*”) and the relevant study design (i.e. treatment or service effectiveness studies). We reviewed the reference lists of key articles and manually added relevant studies missed by the automated search. While we placed no restrictions on study design or data quality, the search did focus on papers that had passed peer-review as a quality control mechanism and excluded grey literature and doctoral dissertations. Language of publication was restricted to English, French, German, Italian or Spanish.

Study eligibility criteria

To be included, studies had to assess the effectiveness or efficacy of an intervention or routine service provision, or report on treatment outcomes as part of exploring moderators of treatment response. Any research design was eligible, including purely qualitative studies. Studies were included if the mean age of the treatment sample was between 12 and 19 years, representing mid to late adolescence as defined by the World Health Organization,[30] and participants had to be either diagnosed with a depressive disorder, help-seeking, or referred for depressive symptoms. Studies focusing on young people with an underlying physical illness, developmental disorder, comorbid psychosis, or personality disorders were excluded. No restrictions were placed on the study setting or type of mental health treatment. We excluded pilot and feasibility studies, as well as studies focusing exclusively on prevention, maintenance, or safety, treatment adherence or engagement.

Study selection and data extraction process

Overall, 7638 titles were identified through the automated and manual searches (see Figure 1) and, after removing duplicates, 7483 were included in the screening of titles and abstracts. Since the search terms were broadly defined to identify any studies relating to internalizing problems, a large number of studies were subsequently excluded as they did not focus specifically on adolescent depression. Likewise, many studies originally identified did not assess treatment effectiveness, but explored the etiology of depression. We retained 320 studies for full-text screening. An independent second reviewer (HB) replicated the full-text screening for 10% of the pre-selected studies, yielding a kappa of 0.68, which indicates a fair to good level of agreement between both reviewers.[31] The final sample consisted of 95 studies, 92 of which were considered for quantitative synthesis, and six for qualitative synthesis, with three studies considered for both. Data was systematically extracted on study characteristics (e.g. study design, participant characteristics, treatment type, date of publication), outcome measures used, and reporters consulted.

Insert Figure 1 here.

Assessment of data quality and risk of bias

The methodological quality of the included studies was assessed using an abbreviated version of the Downs and Black[32] checklist for data quality in randomized and non-randomized healthcare intervention studies. The abbreviated checklist included 23 equally weighted criteria, with the total attainable data quality score being 23. Three purely qualitative studies were excluded from the data quality assessment.

Data synthesis and statistical analysis

We identified outcome domains and subdomains by mapping outcome measures used to the primary concept they were designed to measure. For conceptual guidance we drew on existing taxonomies of treatment outcomes,[10,17,20,34] as well as handbooks of psychometric measures[35] and the original studies reporting on the relevant measure’s development. To ensure the reliability and validity of the coding, a second author (HB) independently coded 25 % of the 121 extracted measures. Both coders (KK and HB) discussed diverging codes and agreed on the final coding. The coding framework thus obtained was also used for a brief thematic analysis[36] of outcomes described in the qualitative studies. Descriptive statistical analysis was performed on the 92 studies that presented quantitative outcome measures (out of the 95 studies included) to explore differences in outcome domains measured and informants consulted over time and in relation to study characteristics. An overview of all included studies can be found in Supplementary Tables 1 and 2 (available online).

Results

Study and participant characteristics

Of the 95 studies under review, the majority (67 %) had been conducted in mental health settings in the United States or Canada; nine studies each had been implemented in the United Kingdom, and in Australia/New Zealand; and the remainder were implemented in Continental Europe, the Middle East and Asia, or multiple regions. Over two thirds of the reviewed studies were randomized control trials (RCTs), followed by observational studies (12 %). The remainder were mixed-methods case studies, retrospective chart reviews and purely qualitative studies. Half of the reviewed studies had the primary objective of assessing efficacy or effectiveness of new or enhanced treatments or service delivery. The other half focused on the assessment of predictors or moderators of treatment response, on cost-effectiveness, or long-term outcomes. These studies often presented secondary analyses of data from large trials such as the Treatment for Adolescents with Depression Study (TADS)[37] or the Treatment of SSRI-resistant Depression in Adolescents (TORDIA) study.[38]

Insert Table 1 here.

The most common type of intervention was the combination of antidepressant medication and psychotherapy (45 %), followed by psychotherapeutic approaches (33 %), medication (18 %), and routine care in the form of treatment as usual, or quality improvement initiatives. The majority of studies (71 %) were conducted in outpatient settings.

The mean data quality score for the 92 quantitative studies based on the modified Downs and Black[32] checklist was 17 (SD = 2.8) out of 23 possible points, with individual scores ranging from 10 to 22. Non-randomized studies (n = 20) had a mean score of 13.7 (SD = 2.0), whereas randomized studies had an average score of 18.1 (SD = 2.2).

The studies involved a total sample of 7394 young people aged 12 to 19 years. The mean age of participants across studies was 15.1 years, 65 % were female, and the majority (82 %) had been recruited based on meeting diagnostic criteria for a depressive disorder, whereas the remainder had been recruited based on a clinical referral or self-referral.

Outcomes measurement across the sample

Across the 92 quantitative studies, a total of 121 outcome measures were identified, which mapped onto 10 overarching outcome domains and 35 sub-domains (see Supplementary Table 3, available online). On average, each study assessed treatment response across 2.1 outcome domains (SD = 2.0) and used four outcome measures, with 14 studies using only a single measure and one study using as many as 14. The domain most frequently measured was that of *depressive symptoms* (see Table 1), assessed by 94 % of studies. Apart from general symptom severity, this domain also included measures assessing suicidal ideation, general psychopathology, and fluctuating mood states associated with depression. Overall, 57 studies used a symptoms measure as their primary outcome measure.

The second most frequently assessed domain was that of *functioning*, measured by 48 studies (52 %), and nominated as a primary outcome by 27. All these 48 studies included a measure of global functioning, two measured the achievement of developmental milestones such as graduating from high school or entering employment, and one measured executive functioning. While some global functioning measures like the Impairment Rating Scale (IRS)[39] and the Behavioral Assessment System for Children-2 (BASC-2)[40] include items related to functioning in school, none of the studies explicitly measured school attendance or performance.

Several studies (13 %) measured not only general depressive symptoms, but also specific *cognitive or behavioral patterns* associated with depression, such as ruminative thinking or negative attributional styles (measured by nine studies), as well as behavioral activation and physical activity (two studies each). Seventeen studies (19 %) measured change in *comorbid symptoms*, such as anxiety, substance misuse, conduct problems, sleeping, compulsions, and disordered eating.

Changes related to any of the other six domains in the coding framework were less frequently measured. The domains of *interpersonal relationships*, *personal growth*, *service satisfaction*, *quality of life*, *parental symptoms*, and *physical health* were each covered by less than one in ten studies. Only three studies measured change in any of these domains as a primary outcome. *Service quality and satisfaction* (n = 8) included measures of consumer satisfaction, as well as therapeutic alliance, treatment adherence, and progress towards patient-defined goals. The domain of *personal growth* (n = 7) covered aspects such as self-concept, coping skills, assertiveness, and autonomy, whereas the domain of *quality of life* (n = 7) captured a broader concept of well-being. Six studies assessed change with regards to *interpersonal relationships*, covering concepts such as social functioning, the functioning of the family system, and peer relationships. Only three and two studies, respectively measured changes in the young person’s *physical health,* or *parental well-being*.

Insert Table 2 here.

Of those six studies that included qualitative analysis of treatment response, all but one measured change in the domain of personal growth, four each covered symptoms, cognitive and behavioral processes, and interpersonal relationships, three covered functioning, two covered comorbid symptoms and one each covered service satisfaction and physical health. The qualitative data touched upon additional subdomains not measured by the quantitative studies, such as school performance and attendance (n = 3), increased motivation and aspirations (n = 2) and calmness (n = 1). They also highlighted a range of additional parent-related outcomes, including parental ability to cope, acceptance of the child's strengths and weaknesses, and parenting practices.

The focus on measuring symptoms and functioning across the reviewed studies is reflected in the fifteen outcome measures most frequently used: all measured either symptoms, functioning, or a combination of both (see Supplementary Table 4, available online). The most commonly used scale was the clinician-rated Children's Depression Rating Scale Revised (CDRS-R),[41] a 17-item interview used to assess depression severity and impairment. The CDRS-R was used by close to half of all quantitative studies (n = 45) and served as a primary outcome measure in 37.

The second most commonly used scale was the clinician-rated Clinical Global Impression Scale Improvement (CGI-I),[42] which assesses the clinician's sense of change in the patient's global functioning.[42] The CGI-I was used by 32 studies and served as a primary outcome measure in 21. The third most commonly used scale was the Beck Depression Inventory (BDI)[31], a youth self-report measure of depressive symptoms used by 16 studies and serving as a primary outcome measure in 5.

Multidimensional outcome measurement by study characteristics

There was an upwards trend in the number of domains covered over time (see Figure 2). Whilst an average of 2.0 (SD = 0.5) domains were covered by studies published in 2007, this had increased to 3.3 (SD = 1.0) by 2015, and to 4.0 (SD = 0.0) in studies published in 2017. In parallel, the number of outcome measures used initially decreased from 3.8 (SD = 1.5) in 2007, to 2.8 (SD = 1.0) in 2013, and has since risen steeply, reaching an average of 8.0 (SD = 1.4) measures in studies published in 2017.

Insert Figure 2 and Figure 3 here side by side.

In terms of the country of implementation, studies conducted in the United Kingdom covered 3.0 (SD = 1.5) domains on average, compared with two in other regions (See Table 2). They used an average of 5.3 (SD = 4.1) measures, compared with 4.2 (SD = 2.5) in the Middle East and East Asia, 3.9 (SD = 2.6) in Australia and New Zealand, 3.5 (SD = 2.4) in North America and 3.1 (SD = 2.0) in Continental Europe.

Regarding study size, small studies (i.e. with sample sizes of 11-50) used the largest average number of individual measures (5.5 measures, SD = 3.3), whereas the most highly powered studies used fewer measures (2.9 measures, SD = 1.6). Likewise, the number of domains covered decreased with study size, from 3.0 domains (SD = 1.6) for studies with up to 10 participants, to 1.9 domains (SD = 0.7) for studies with more than 300 participants. Case study and qualitative designs covered a larger average number of outcome domains than studies with other designs. The number of domains covered and of measures did not vary considerably depending on the data quality score.

Informants consulted across outcome domains

Thirty-nine studies (42 %) assessed change in outcomes using exclusively clinician-rated measures, whereas 30 studies (33 %) used both clinician and youth-reported measures, ten (11 %) used youth and parent self-report, and eight (9 %) relied entirely on youth report. Overall, youth report was considered by more than half of the studies (53 %). However, primary outcome measures, as identified by the relevant study authors, predominantly relied on clinician-report (75 %).

As shown in Figure 3, there was an inconsistent trend in the use of youth self-report over time. Of the seven studies published in 2007, 57 % used at least one youth self-report measure and the percentage remained relatively stable until 2012, then dropping to zero in 2013 (n = 4) and remaining low at 14 % for the seven studies published in 2014. In contrast, all fourteen studies published since 2015 have included a youth self-report measure.

In some outcome domains, youth self-report served as the primary source of information, whereas other domains were mainly assessed using clinician report. All relevant studies consulted young people to measure change in cognition and behavior, and quality of life, with youth being the only relevant informant in 80 % of these studies. Likewise, youth self-report was the main source of information concerning the domains of personal growth, service satisfaction, and interpersonal relationships. Clinician-report dominated the measurement of depressive symptoms and functioning. While youth report on depressive symptoms was gathered by half of all studies, only six percent included youth-self report for functioning.

Discussion

This review finds that the measurement of treatment effectiveness for adolescent depression over the past decade has focused on symptoms and functioning. Other domains have occasionally been covered as secondary outcomes, using a plethora of different measures (see Supplementary Table 3, available online), generally without providing a detailed rationale for their choice. Clinician-report is the predominant source of information for primary outcomes. Over half of the reviewed studies included youth self-report, but mainly to assess secondary outcomes. We observe a trend towards more multidimensional measurement and the inclusion of youth self-report in most recently published studies.

The predominance of symptoms and functioning measurement relative to other domains is in line with findings from earlier reviews of treatment outcomes in child mental health across disorders.[9,10,29,44,45] This suggests a high degree of continuity in the predominance of unidimensional outcomes measurement over the past decades, with change only showing for the most recent studies. This recent increase may reflect the growing policy interest in outcomes measurement and a parallel growth in research about patient-reported and goal-based outcomes.[26,46–48] Two recent taxonomies have emphasised the importance of measuring domains beyond symptoms,[17,20] as have a small number of qualitative studies.[49,50] However, with the exception of one paper[51] none of the reviewed studies referred to this emerging literature or discussed their measurement approach from a multidimensional perspective.

The predominance of clinician-report in the measurement of primary outcomes may reflect a belief that clinicians provide a global view of clinical improvement, as their assessments often consider information provided by young people and their parents.[52] Yet, the evidence suggests that clinicians accord greater importance to parent report than to youth self-report,[53–55] and that different informants often provide inconsistent but valid accounts.[56,57] Discrepancies between these accounts may suggest that treatment has altered a young person's behavior in one context but not another, and are thus highly informative[57] Yet, the reviewed studies did not explore such differences.

The recent rise in studies that include youth self-report may reflect a move towards more patient-centered and youth-guided care as underpinning principles for practitioners and funders.[58,59] For instance, the US Substance Abuse and Mental Health Services Administration (SAMHSA) mandates that young people are systematically consulted on the design, implementation and evaluation of community-based systems of care funded through its Children's Mental Health Initiative (CMHI).[60] Nevertheless, the reviewed studies rarely elaborated on their rationale for including youth self-report and did not refer to the youth advocacy movement spearheaded by SAMHSA, or comparable initiatives.

The relative absence of multidimensional measurement has theoretical and practical implications. Arguably, measurement approaches should provide the best possible fit to grasp the theoretical concept that is the disorder, as well as conceptualizations of how treatment works.[61] Depression is multifactorial in terms of its risk factors[62–64] and its pervasive impact on psychosocial functioning and well-being.[65] Evidence suggests that symptoms change does not reliably predict change in other domains.[28,44,66,67] It appears crucial to assess outcomes across a range of domains to gain a nuanced understanding of whether, how, and when treatment is effective, and to avoid premature conclusions that may have significant impact on funding and policy decisions.

Using multiple outcome measures and informants is likely to be challenging in practice. Young people are most willing to engage if the changes measured are meaningful to them.[20,68,69] Clinicians require measures that deliver actionable results such as alerting them to warning signs for individual patients, whereas commissioners may prefer data that can support benchmarking.[70] There may be inherent tension between these needs. A mitigation strategy may be to develop flexible core measure sets that include those measures needed for benchmarking, while recommending measures within other domains that could be integrated, based on shared decision-making with parents and young people. Such a multi-dimensional core battery would also benefit clinical research, enhancing the comparability of findings across studies and addressing the current fragmentation of measures.

Policy makers have begun to encourage or mandate the use of common sets of measures, such as the Core Set of Children’s Health Care Quality Measures for Medicaid and the Children's Health Insurance Program (CHIP) in the US[71] or the Australian Government’s National Outcomes and Casemix Collection.[5] In the summer of 2018, the US National

Committee for Quality Assurance (NCQA) launched a National Collaborative for Innovation in Quality Measurement (NCINQ), which will convene a national panel of advisors including youth and family representatives to support health plans and states in researching possible improvements to depression care for adolescents using three core quality measures included in the NCQA's Healthcare Effectiveness Data and Information Set (HEDIS).[72] In light of such significant national attention to quality and outcomes measurement for adolescent depression, clarifying what is important to measure appears both imperative and urgent.

Several limitations of this study should be noted. First, while the language of publication was not restricted to English, no foreign-language study met the inclusion criteria, and around half of the included studies were conducted in the United States. Therefore, regional differences in outcomes measurement could not be explored in detail. Second, the systematic literature search only considered studies published in peer-reviewed journals and may have missed reports of program or service evaluations published elsewhere. Studies conducted in routine care settings may therefore be under-represented relative to more rigorously designed clinical trials that passed peer review. Additional research is needed to explore what outcome domains are measured in real-world CAMH settings, and the extent to which the picture converges with that found by this review.

Third, it is worth highlighting that several of the reviewed studies link to larger clinical trials. For instance, 17 studies used data from the TADS[37] and 14 studies used data from the TORDIA[38] study. Even though each of these studies answered different research questions and used different combinations of measures, all used the same primary outcome measures, which means that the relevant studies in our review sample are not entirely independent from one another.

As this review demonstrates, the recent treatment effectiveness literature for adolescent depression focuses mainly on symptoms and functioning as reported by clinicians, at the expense of exploring change across a broader range of domains and informants. The current momentum around patient-centred and youth-guided care provides a historical chance to accelerate multi-dimensional and multi-informant measurement and to reconsider what is important to measure for different stakeholders, including young people themselves. At a time where there is more scrutiny on treatment outcomes than ever before, making sure that we measure what matters most to patients is paramount.

References

1. Kleinert S, Horton R. From universal health coverage to right care for health. *Lancet*. 2017;390(10090):101-102.
2. Gray M. Value based healthcare. *BMJ*. 2017;356:1-3.
3. Chee TT, Ryan AM, Wasfy JH, Borden WB. Current State of Value-Based Purchasing Programs. *Circulation*. 2016;133(22):2197-2205.
4. Department of Health. *Talking Therapies: A Four-Year Plan of Action. A Supporting Document to No Health without Mental Health: A Cross- Government Mental Health Outcomes Strategy for People of All Ages*.; 2011. http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_123759. Accessed July 3, 2018.
5. Department of Health and Ageing. *Mental Health National Outcomes and Casemix Collection: Overview of Clinician-Rated and Consumer Self-Report Measures V1.50*. Canberra; 2003. https://www.amhocn.org/sites/default/files/publication_files/mhnocc_overview_doha.pdf. Accessed July 3, 2018.
6. National Quality Forum. *Measure Applications Partnership : Strengthening the Core Set of Healthcare Quality Measures for Children Enrolled in Medicaid and CHIP , 2017*.; 2017. https://www.qualityforum.org/Publications/2017/08/Strengthening_the_Core_Set_of_Healthcare_Quality_Measures_for_Children_Enrolled_in_Medicaid,_2017.aspx. Accessed July 3, 2018.
7. World Health Organization. *Depression and Other Common Mental Disorders: Global Health Estimates*. Geneva; 2017.

http://www.who.int/mental_health/management/depression/prevalence_global_health_estimates/en/. Accessed July 3, 2018.

8. National Committee for Quality Assurance. National Collaborative for Innovation in Quality Measurement : Implementing and Improving. 2018. http://www.ncqa.org/Portals/0/HEDISQM/LearningCollaborative/NCINQ_Depression_Quality_Improvement_Collaborative.pdf?ver=2018-04-11-222503-473. Accessed July 3, 2018.

9. Jensen PS, Hoagwood K, Petti T. Outcomes of mental health care for children and adolescents: II. Literature review and application of a comprehensive model. *J Am Acad Child Adolesc Psychiatry*. 1996;35(8):1064-1077.

10. Hoagwood KE, Jensen PS, Acri MC, Serene Olin S, Eric Lewandowski R, Herman RJ. Outcome domains in child mental health research since 1996: Have they changed and why does it matter? *J Am Acad Child Adolesc Psychiatry*. 2012;51(12):1241-1260.

11. Lambert MJ, Whipple JL, Hawkins EJ, Vermeersch DA, Nielsen SL, Smart DW. Is it time for clinicians to routinely track patient outcome? A meta-analysis. *Clin Psychol Sci Pract*. 2003;10(3):288-301.

12. Marks S. Ecology, Humanism and Mental Health in Communist Czechoslovakia. In: Marks S, Savelli M, eds. *Psychiatry in Communist Europe*. London: Palgrave; 2015.

13. Ludvigsen K, Seip AA. The establishing of Norwegian child psychiatry: ideas, pioneers and institutions. *Hist Psychiatry*. 2009;20(77 Pt 1):5-26.

14. Binder P, Holgersen H, Nielsen GH. What is a “ good outcome ” in psychotherapy ? A qualitative exploration of former patients ’ point of view. *Psychother Res*. 2010;20(3):285-294.

15. Miller G. Winifred Rushforth and the Davidson Clinic for Medical Psychotherapy: a

case study in the overlap of psychotherapy, Christianity and New Age spirituality. *Hist Psychiatry*. 2015;26(3):303-317.

16. Harding C. Japanese psychoanalysis and Buddhism: the making of a relationship. *Hist Psychiatry*. 2014;25(2):154-170.

17. Bradley J, Murphy S, Fugard AJB, Nolas S-M, Law D. What kinds of goals do children and young people set for themselves in therapy? Developing a goals framework using CORC data. *Child Fam Clin Psychol Rev*. 2013;1(1):8-18.

18. Yeh M, Weisz JR. Why are we here at the clinic? Parent-child (dis)agreement on referral problems at outpatient treatment entry. *J Consult Clin Psychol*. 2001;69(6):1018-1025.

19. Lelliott P, Beevor A, Hogman G, Hyslop J, Lathlean J, Ward M. Carers' and Users' Expectations of Services--User version (CUES-U): a new instrument to measure the experience of users of mental health services. *Br J Psychiatry*. 2001;179:67-72.

20. Childs J, Deighton J, Wolpert M. *Defining and Measuring Mental Health and Wellbeing: A Response Mode Report Requested by the Department of Health for the Policy Research Unit in the Health of Children, Young People and Families*. London; 2013. https://www.ucl.ac.uk/evidence-based-practice-unit/sites/evidence-based-practice-unit/files/pub_and_resources_project_reports_defining_and_measuring_mental_health.pdf. Accessed July 3, 2018.

21. Leamy M, Bird V, Le BC, Williams J, Slade M. Conceptual framework for personal recovery in mental health: systematic review and narrative synthesis. *BrJ Psychiatry*. 2011;199(1472-1465 (Electronic)):445-452.

22. 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):1055-1063.

23. Weisz JR, Kuppens S, Ng MY, 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. *Am Psychol*. 2017;72(2):79-117.

24. Department of Health. *Future in Mind: Promoting, Protecting and Improving Our Children and Young People's Mental Health and Wellbeing*. NHS England Publication Gateway Ref. No 02939.; 2015. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/414024/Childrens_Mental_Health.pdf. Accessed July 3, 2018.

25. De Los Reyes A, Augustein T, Wang M, et al. The Validity of the Multi-Informant Approach to Assessing Child and Adolescent Mental Health. *Psychol Bull*. 2015;141(4):858-900.

26. Coulter A. Measuring what matters to patients. *BMJ*. 2017;356:j816.

27. Mulley A, Coulter A, Wolpert M, Richards T, Abbasi K. New approaches to measurement and management for high integrity health systems. *BMJ*. 2017;356:j1401.

28. Becker KD, Brandt NE, Stephan SH, Chorpita BF. A review of educational outcomes in the children's mental health treatment literature. *Adv Sch Ment Health Promot*. 2014;7(1):5-23.

29. Weisz JR, Doss AJ, Hawley KM. Youth psychotherapy outcome research: a review and critique of the evidence base. *Annu Rev Psychol*. 2005;56:337-363.

30. World Health Organisation. Recognizing adolescence. 2017. http://apps.who.int/adolescent/second-decade/section2/page1/recognizing-

adolescence.html. Accessed March 1, 2018.

31. Fleiss JL. *Statistical Methods for Rates and Proportions.* New York: John Wiley and Sons; 1981.

32. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health*. 1998;52(6):377-384.

33. Moher D, Liberati A, Tetzlaff J, Altman D, The PRISMA Group. Preferred Reporting Items for Systematic Reviews and MetaAnalyses: The PRISMA Statement. *PLOS Med*. 2009;6(7):e1000097.

34. Grosse M, Grawe K. BERN INVENTORY OF TREATMENT GOALS: PART 1. Development and First Application of a Taxonomy of Treatment Goal Themes. *Psychother Res*. 2002;12(1):79-99.

35. Rush AJ, First MB, Blacker D. *Handbook of Psychiatric Measures*. 2nd editio. Washington DC: American Psychiatric Association Publishing; 2008.

36. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol*. 2006;3(2):77-101.

37. Treatment for Adolescents With Depression Study (TADS) Team. Fluoxetine, Cognitive-Behavioral Therapy, and Their Combination for Adolescents With Depression. *Jama*. 2004;292(7):807.

38. Brent D, Emslie G, Clarke G, et al. Switching to another SSRI or to venlafaxine with or without cognitive behavioral therapy for adolescents with SSRI-resistant depression: The TORDIA randomized controlled trial. *JAMA - J Am Med Assoc*. 2008;299(8):901-913.

39. Fabiano GA, Pelham, Jr. WE, Waschbusch DA, 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):369-385.

40. Reynolds CR. Behavior Assessment System for Children. In: Weiner IB, Edward CW, eds. *The Corsini Encyclopedia of Psychology*. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2010.

41. Poznanski E, Mokros H. *Children's Depression Rating Scale–Revised (CDRS-R)*. Los Angeles: Western Psychological Services; 1996.

42. Busner J, Targum SD. The clinical global impressions scale: applying a research tool in clinical practice. *Psychiatry (Edgmont)*. 2007;4(7):28-37.

43. 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.

44. Becker KD, Chorpita BF, Daleiden EL. Improvement in symptoms versus functioning: How do our best treatments measure up? *Adm Policy Ment Heal Ment Heal Serv Res*. 2011;38(6):440-458.

45. Weisz JR, Doss AJ, Hawley KM. Youth Psychotherapy Outcome Research: A Review and Critique of the Evidence Base. *Annu Rev Psychol*. 2005;56(1):337-363.

46. Jacob J, Edbrooke-Childs J, Law D, Wolpert M. Measuring what matters to patients: Using goal content to inform measure choice and development. *Clin Child Psychol Psychiatry*. 2017;22(2):170-186.

47. Wolpert M, Ford T, Trustam E, et al. Patient-reported outcomes in child and adolescent mental health services (CAMHS): Use of idiographic and standardized measures. *J Ment Heal*. 2012;21(2):165-173.

48. Edbrooke-Childs J, Wolpert M, Deighton J. Introduction to the special section on implementing feedback from outcome measures in child and adult mental health services. Bickman Carlier, Fleming, Gleacher, Knaup B, ed. *Adm Policy Ment Heal Ment Heal Serv Res*. 2016;43(3):277-278.

49. Moltu C, Stefansen J, Nøtnes JC, Skjølberg Å, Veseth M. What are "good outcomes" in public mental health settings? A qualitative exploration of clients' and therapists' experiences. *Int J Ment Health Syst*. 2017;11(1):12.

50. Lavik KO, Veseth M, Frøysa H, Binder PE, Moltu C. What are " good outcomes " for adolescents in public mental health settings ? *Int J Ment Health Syst*. 2018;12(3):1-10.

51. Weisz JR, Southam-Gerow MA, Gordis EB, et al. Cognitive-Behavioral Therapy Versus Usual Clinical Care for Youth Depression: An Initial Test of Transportability to Community Clinics and Clinicians. *J Consult Clin Psychol*. 2009;77(3):383-396.

52. De Los Reyes A, Kundey SMA, Wang M. The end of the primary outcome measure: A research agenda for constructing its replacement. *Clin Psychol Rev*. 2011;31(5):829-838.

53. Loeber R, Green S, Lahey B. Mental Health Professionals' Perception of the Utility of Children, Mothers, and Teachers as Informants on Childhood Psychopathology. *J Clin Child Adolesc Psychol*. 1990;19(2):136-143.

54. Hawley KM, Weisz JR. Child , Parent , and Therapist (Dis) Agreement on Target Problems in Outpatient Therapy : The Therapist ' s Dilemma and Its Implications. 2003;71(1):62-70.

55. Kramer T, Phillips SD, Hargis MB, Miller TL, Burns BJ, Robbins JM. Disagreement between parent and adolescent reports of functional impairment. *J Child Psychol Psychiatry Allied Discip*. 2004;45(2):248-259.

56. De Reyes AL, Kazdin AE. Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychol Bull*. 2005;131(4):483-509.

57. Hunsley J, Mash EJ. Evidence-Based Assessment. *Annu Rev Clin Psychol*. 2007;3(1):29-51.

58. Department of Health. *Equity and Excellence: Liberating the NHS*. London; 2010. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/213823/dh_117794.pdf. Accessed July 3, 2018.

59. National Quality Forum. Shared decision making: A standard of care for all patients. A National Quality Partners ™ Action Brief. 2017. http://www.qualityforum.org/Publications/2017/10/NQP_Shared_Decision_Making_Action_Brief.aspx. Accessed July 3, 2018.

60. Davis-Brown K, Carter N, Miller BD. Youth Advisors Driving Action: Hearing the Youth Voice in Mental Health Systems of Care. *J Psychosoc Nurs Ment Health Serv*. 2012;50(3):39-43.

61. Dirks MA, De Los Reyes A, Briggs-Gowan M, Cella D, Wakschlag LS. Annual research review: Embracing not erasing contextual variability in children’s behavior - Theory and utility in the selection and use of methods and informants in developmental psychopathology. *J Child Psychol Psychiatry Allied Discip*. 2012;53(5):558-574.

62. Clarke GN, DeBar LL. Group Cognitive-Behavioral Treatment for Adolescent Depression. In: Weisz JR, Kazdin AE, eds. *Evidence-Based Psychotherapies for Children and Adolescents.* 2nd editio. New York: The Guilford Press; 2010:110-125.

63. Kassis W, Artz S, White J. Understanding Depression in Adolescents: A Dynamic

Psychosocial Web of Risk and Protective Factors. *Child Youth Care Forum*. 2017;46(5):721-743.

64. Weeks M, Ploubidis GB, Cairney J, Wild TC, Naicker K, Colman I. Developmental pathways linking childhood and adolescent internalizing, externalizing, academic competence, and adolescent depression. *J Adolesc*. 2016;51:30-40.

65. Klein DN, Dougherty LR, Olino TM. Toward Guidelines for Evidence-Based Assessment of Depression in Children and Adolescents. *J Clin Child Adolesc Psychol*. 2005;34(3):412-432.

66. Karpenko V, Owens JS. Adolescent Psychotherapy Outcomes in Community Mental Health: How Do Symptoms Align with Target Complaints and Perceived Change? *Community Ment Health J*. 2013;49(5):540-552.

67. Brookman-Frazee L, Haine RA, Garland AF. Measuring Outcomes of Real- World Youth Psychotherapy : Whom to Ask and What to Ask ? *Psychiatr Serv*. 2006;57(10):1373-1375.

68. Moran P, Kelesidi K, Guglani S, Davidson S, Ford T. What do parents and carers think about routine outcome measures and their use? A focus group study of CAMHS attenders. *Clin Child Psychol Psychiatry*. 2012;17(1):65-79.

69. Stasiak K, Parkin A, Seymour F, et al. Measuring outcome in child and adolescent mental health services: Consumers’ views of measures. *Clin Child Psychol Psychiatry*. 2013;18(4):519-535.

70. 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.

71. Measure Applications Partnership. *Strengthening the Core Set of Healthcare Quality*

Measures for Adults Enrolled in Medicaid, 2017. Washington, DC; 2017. https://www.qualityforum.org/Publications/2017/08/Strengthening_the_Core_Set_of_Healthcare_Quality_Measures_for_Adults_Enrolled_in_Medicaid,_2017.aspx. Accessed July 3, 2018.

72. National Committee for Quality Assurance (NCQA). National Collaborative for Innovation in Quality Measurement: Implementing and Improving. http://www.ncqa.org/hedis-quality-measurement/research/ncinq. Published 2018. Accessed July 3, 2018.

73. Wachs TD, Rahman A. The nature and impact of risk and protective influences on children's development in low-income countries. P. R. Britto, P. L. Engle & CMS, ed. *Handb early Child Dev Res its Impact Glob policy*. 2013:85-122.

74. Weisz JR, Chorpita BF, Frye A, et al. Youth Top Problems: using idiographic, consumer-guided assessment to identify treatment needs and to track change during psychotherapy. *J Consult Clin Psychol*. 2011;79(3):369-380.

Table 1: Outcome Domains and Subdomains by Frequency of Measurement

Domains and subdomains	Number (and %) of studies measuring this domain in general (n = 92)	 as primary outcome (n = 68)	Measured in qualitative studies (n = 6)	Included in existing taxonomies
Symptoms	**86 (93.5)**	**57 (83.8)**		
Depressive symptoms and diagnoses	86 (93.5)	57 (83.8)	✓	✓[17,20,22]
Suicidality	15 (16.3)	3 (4.4)	✓	✓[34]
Self-harm	1 (1.1)		✓	
Mood states / distress / anger	2 (2.2)			
General psychopathology	6 (6.5)			
Comorbid symptoms	**17 (18.5)**	**3 (4.4)**	✓	✓[17,20,22,34]
Anxiety	7 (7.6)	1 (1.5)		
Alcohol and drug misuse	6 (6.5)	1 (1.5)	✓	
Conduct problems	5 (5.4)	1 (1.5)		
Disordered eating	1 (1.1)			
Obsessive-compulsive disorder	1 (1.1)			
Sleeping problems	2 (2.2)		✓	
Non-specific comorbidities	2 (2.2)			
			✓	
Cognitive and behavioral processes linked to depression	**12 (13)**	**2 (2.9)**	✓	
Cognitive processes	9 (9.8)		✓	
Behavioral Activation	2 (2.2)	2 (2.9)	✓	
Physical activity and exercise	2 (2.2)			✓[34]
Functioning	**48 (52.2)**	**27 (39.7)**	✓	
Academic achievement			✓	✓[17,20,22]
Academic behavior and skills				✓[28]
Academic self-efficacy				✓[28]
Communication skills			✓	✓[44]
Executive functioning	1 (1.1)		✓	✓[20]
Global functioning	48 (52.2)	27 (39.7)	✓	✓[17,20,22]
Life course achievements	2 (2.2)			
Interpersonal relationships	**6 (6.5)**		✓	
Attachment style	1 (1.1)			
Family functioning and relations	4 (4.3)		✓	✓[17,20,22]
Peer relationships	1 (1.1)		✓	✓[20,73]
Social functioning	3 (3.3)		✓	✓[20,22,44]
Taking on home chores			✓	✓[17]
Personal growth	**7 (7.6)**	**1 (1.5)**	✓	
Assertiveness	1 (1.1)		✓	✓[34]
Attitudes towards self (eg, self-esteem)	1 (1.1)		✓	✓[17,34]
Autonomy	1 (1.1)		✓	✓[17,20]
Balance, calmness and composure			✓	✓[34]
Coping with mood, thoughts and feelings	2 (2.2)		✓	✓[17,20]
Empowerment and self-efficacy			✓	✓[20]
Motivation / aspirations			✓	
Locus of control				✓[20]
Processing personal history				✓[17,20]

Service use and satisfaction	**8 (8.7)**		✓	
Client satisfaction	3 (3.3)			✓[20,22]
Idiographic goals	2 (2.2)		✓	✓[74]
Use of other services	4 (4.3)			
Therapeutic alliance	2 (2.2)			
Treatment retention	1 (1.1)			
Retention of therapeutic content	1 (1.1)			
Quality of life	**7 (7.6)**			
Health-related quality of life	7 (7.6)			✓[17,22]
Physical health	**3 (3.3)**	**2 (2.9)**	✓	
Physical health			✓	✓[10]
Biomarkers	3 (3.3)	2 (2.9)		✓[10]
Parental symptoms	**2 (2.2)**		✓	
Parental symptoms	2 (2.2)			✓[10,20,22]
Parenting practices			✓	
Parental coping			✓	

Table 2: Study Characteristics, Number of Outcome Domains Covered, and Measures Used

Study characteristics	n (%)	Average number of domains covered	Average number of measured used
Publication year			
2007 - 2012	68 (71.6)	1.9	3.2
2013 - 2017	27 (28.4)	2.7	4.7
Country			
United States and Canada	62 (67.4)	2.0	3.5
United Kingdom	9 (9.5)	3.0	5.3
Continental Europe	7 (7.4)	2.0	3.1
Australia and New Zealand	9 (9.5)	2.2	3.9
Middle East and Asia	6 (6.3)	2.0	4.2
Multiple regions	2 (2.1)	2.0	2.0
Study type			
Randomized controlled trial	69 (72.6)	2.1	3.6
Observational study	11 (11.6)	2.2	4.3
Naturalistic follow-up	5 (5.3)	1.4	2.4
Retrospective chart review	3 (3.2)	1.3	2.0
Single case study	4 (4.4)	3.5	5.3
Qualitative study	3 (3.2)	4.0	—
Study objective			
Assess treatment effectiveness	49 (51.6)	2.5	4.5
Study predictors / mediation factors	33 (34.8)	1.7	2.4
Examine long-term outcomes	6 (6.3)	1.7	3.7
Cost-effectiveness	4 (4.2)	2.0	2.0
Other[b]	3 (3.2)	3.0	6.0
Sample size			
1–10	7 (7.4)	3.0	4.4
11–50	14 (14.7)	2.6	5.5
51–100	15 (15.8)	2.2	4.3
101–300	23 (24.2)	2.0	3.1
301–500	36 (37.9)	1.9	2.9
Study intervention			
Combined treatment	41 (43.2)	1.9	3.1
Psychotherapy	31 (32.6)	2.7	4.9
Medication	17 (17.9)	1.9	3.3
Routine care	6 (3.2)	2.2	1.8
Data quality assessment scores			
Quartile 1 (10-15)	23 (25.3)	2.1	3.5
Quartile 2 (15-18)	23 (25.3)	2.2	3.9
Quartile 3 (18-19)	23 (25.3)	1.9	3
Quartile 4 (19-22)	22 (24.2)	2.3	3.8

Note:[a] Includes one study conducted in the United States, Canada, Mexico, and Argentina; and one study conducted in the United States, Eastern Europe, South Africa, and Western Europe.

[b] "Other" objectives include: Assessing the transportability of an intervention, reflecting on lessons learned, and assessing treatment experience qualitatively.

Figure 1: Flow Diagram of the Screening Process, Adapted From Moher *et al*.[33]

Figure 2: Average Number of Domains Assessed and Measures Used by Year of Publication (2007 – 2017)

Figure 3: Percentage of Studies Including Youth Self-Report by Year of Publication (2007 – 2017)