

Selection of a Child Clinical Outcome Measure for Statewide Use in Publicly Funded Outpatient Mental Health Programs

Bonnie T. Zima, M.D., M.P.H., F. Alethea Marti, Ph.D., Christopher E. Lee, M.P.H., Nadereh Pourat, Ph.D.

Objective: This study describes the process of choosing a clinical outcome measure for a statewide performance outcome system for children receiving publicly funded mental health services in California.

Methods: The recommendation is based on a five-phase approach, including an environmental scan of measures used by state mental health agencies; a statewide provider survey; a scientific literature review; a modified Delphi panel; and final rating of candidate measures by using nine minimum criteria informed by stakeholder priorities, scientific evidence, and state statute.

Results: Only 10 states reported use of at least one standardized measure for outcome measurement. In California, the most frequently reported measures were the Child and Adolescent Needs and Strengths (CANS) (N=33), the Child Behavior Checklist (N=14), and the Eyberg Child Behavior Inventory (N=12). Based on modified Delphi panel ratings,

only the Achenbach System of Empirically Based Assessment, the Strengths and Difficulties Questionnaire, and the Pediatric Symptom Checklist (PSC) were rated on average in the high-equivocal to high range on effective care, scientific acceptability, usability, feasibility, and overall utility. The PSC met all nine minimum criteria for recommendation for statewide use. In its final decision, the California Department of Health Care Services mandated use of the PSC and CANS.

Conclusions: There is a lack of capacity to compare child clinical outcomes across states and California counties. Frequently used outcome measures were often not supported by scientific evidence or Delphi panel ratings. Policy action is needed to promote the selection of a common clinical outcome measure and measurement methodology for children receiving publicly funded mental health care.

Psychiatric Services 2019; 70:381–388; doi: 10.1176/appi.ps.201800424

Providing high-quality care at lower costs is a national goal (1, 2). If it is to be achieved, the primary driver is envisioned to be quality measurement of clinical outcomes that are aligned with financial incentives (3–6). Quality measurement of mental health care, however, has lagged behind advances in other health care sectors, with disproportionately less attention paid to child mental health and outcomes (7–13). In 2013, only 29 state Medicaid behavioral health agencies provided online information related to measuring behavioral health care quality; use of quality measures varied widely by state, with very few targeting care for children (14). Among the 26 measures in the 2018 core set of children's health care quality measures for Medicaid and the Children's Health Insurance Program, only four are related to behavioral health (15), and the relationship between adherence and improved clinical outcomes has not been established (10, 13, 16, 17).

As early as 1991, several legislative mandates in California called for the development of a statewide system for publicly reporting the quality of mental health care and its outcomes

over time. This requirement is embedded within a series of laws that seek to stabilize funding for community mental

HIGHLIGHTS

- Only one out of five state mental health agency Web sites reported use of any standardized measure to track clinical outcomes for children.
- At the state and California county levels, measures varied widely, making it impossible to compare child clinical outcomes across states and across counties within California.
- Following a five-phase approach, the Pediatric Symptom Checklist was recommended for statewide use to track clinical outcomes among children receiving publicly funded community-based mental health services in California.
- Policy action is needed to promote the selection of a common clinical outcome measure for children and to standardize measurement methods.

health programs by shifting administrative and financial responsibility to county mental health agencies and earmarking specific tax revenues for mental health care (18–20). For children, early efforts to measure performance included documenting high need for mental health care in select county programs (21, 22), assessing agreement between child functional measures (23), and describing foster home and state hospital utilization and expenditures among counties implementing system-of-care principles (24).

In 2012, the legislative mandate to transfer the administration of all Medicaid-funded mental health services to the state of California's Department of Health Care Services (DHCS) was amended to include a statute to "develop a performance outcome system for early and periodic screening, diagnosis, and treatment mental health services that will improve outcomes at the individual and system levels and will inform fiscal decision making related to the purchase of services" (25). Yet, despite these policies, there remains a need to develop a robust data infrastructure for quality monitoring and a standardized approach for measuring child outcomes (19, 26–28).

In this context, DHCS contracted with a major university to address the question, "What is the best statewide approach to evaluate functional status for children and youth that are served by the California public specialty mental health service system?" (29). As the recipients of this contract, we sought to recommend a standardized child measure of functioning for statewide use. To do so, we used a five-phase approach consisting of an Internet environmental scan of measures used by state mental health agencies; a statewide provider survey; a scientific literature review; a modified Delphi panel; and final ratings of candidate measures on the basis of nine minimum criteria informed by stakeholder priorities, scientific evidence, and the performance outcome system statute. At the conclusion of the project, we prepared a report to the state outlining our recommendations for a statewide performance outcome measurement system (30). This article builds on that report by providing a fuller examination of the modified Delphi panel ratings, using qualitative data to explain and identify stakeholder priorities. The article also discusses the final recommendation and implementation plan from the DHCS mental health services division (DHCS-MHSD) and briefly summarizes the study's methods and main findings.

METHODS

Identification of Candidate Measures

To identify a pool of candidate measures, we conducted an environmental scan, a statewide provider survey, and a scientific literature review. The environmental scan examined mental health agency Web sites in 49 states (excluding California) to identify which states used standardized measures to screen for mental health service need or track clinical outcomes for children served by publicly funded specialty mental health programs from December 2015

through February 2016. In addition, a statewide provider survey was conducted by using Survey Monkey in December 2015 to identify which standardized measures of child functioning were used in community-based mental health programs within California and how they were used. The provider sample included behavioral health directors or their designee in 56 of the state's 58 (97%) counties. Exploratory findings from a purposive sample of 21 contracted providers are not reported.

Further, a comprehensive scientific literature scan was conducted by using SCOPUS, PubMed, and PsycINFO to identify peer-reviewed studies from the previous 5 years (2010–2015) that used standardized measures to track clinical outcomes for children ages 0 to 18 who were receiving community-based, outpatient mental health services. Eligibility criteria were peer-reviewed articles published between 2010 and 2015, English-language abstracts, and use of at least one standardized measure that compares change in the child's symptoms or functioning across at least two time points. The scan excluded studies with target populations that did not meet medical necessity criteria for Medicaid reimbursement in California's publicly funded specialty mental health outpatient programs (e.g., primary diagnosis of drug, alcohol, or tobacco use disorder or neurodevelopmental delay).

The final list of candidate measures was merged from these three data sources. Eligibility criteria included use by one state mental health agency, use by two or more California county mental health agencies, or having been used as a clinical outcome measure in three published studies from the literature review. Proprietary and publicly available measures were included. The list did not include measures designed to track individualized outcomes (e.g., therapy progress using more restricted age- or disorder-specific measures or treatment plan goals) because they are not suitable for assessing the effectiveness of care at an aggregate level (provider, program, or county).

Modified Delphi Panel

The modified Delphi method, also called the RAND/University of California at Los Angeles (UCLA) appropriateness method, is a well-established approach that combines scientific evidence and judgment to produce the best possible information (31). The original method entails assessment of existing scientific evidence by a group of nine medical experts, anonymous ranking of quality indicators based on scientific evidence and expert opinion, confidential feedback to panel members on their responses in relation to the rest of the group, and a discussion among the panel followed by a confidential final ranking of the quality indicators (32).

For this project, the method was adapted to expand the breadth of expertise by using a 14-member panel and add ratings for scientific acceptability, feasibility, and usability by using criteria from the National Quality Forum (33). Panel members were purposively selected by using a partnered approach to include expertise in the delivery of publicly

funded child mental health care from a variety of perspectives as well as to include participants from urban and rural counties. Each panelist received a manual containing a summary of features (description, logistics, psychometric properties, and strength of evidence) and scientific evidence tables for each of the candidate measures (30). The strength of the evidence for use as an outcome measure in community-based child mental health programs was rated by using the Oxford Centre for Evidence-Based Medicine (CEBM) levels of evidence. The CEBM protocol ranks the strength of evidence, based on study design and methodologic rigor, from level 1, individual randomized clinical trials with more than 80% follow-up, to level 5, expert opinion or inconclusive evidence (13, 34). For all 11 candidate measures, the strength of evidence for the outcome studies was critically reviewed and assigned a ranking by a board-certified child psychiatrist.

Using a 9-point Likert scale, panelists were also asked to rate the measures on four domains (1, lowest; 4–6, equivocal; and 9, highest) and overall utility (1, definitely would not recommend; 4–6, equivocal; and 9, would definitely recommend). The domains were marker of effective care (the extent to which improvement in the outcome, as assessed by this measure, is an indicator of effective care), scientific acceptability (the extent to which published scientific evidence supports the use of the measure for tracking clinical outcomes in community-based mental health programs, including three subdomains—reliability, validity, and strength of evidence), usability (the extent to which the intended audience can understand the measure's scores and find them useful for decision making), and feasibility (the extent to which data obtained from the measure are readily available or can be captured without undue burden—i.e., no formal training required—and could be implemented by counties to track clinical outcomes). Overall utility was defined as the extent to which a panelist would recommend it for statewide use to track clinical outcomes among children and youths served in publicly funded and community-based specialty mental health programs. Following discussion, panelists confidentially re-rated the measure.

To enrich findings from the panel ratings, the discussion was audio-taped and transcribed for qualitative analysis (35). Transcripts were coded by topic, with inductive codes for theme, affect (positive or negative), and the four specified domains (36). Each measure's discussion was analyzed independently and condensed into a synthesis of topics (e.g., features and specific concerns). The full session was then analyzed holistically into a synthesis of common themes appearing repeatedly across multiple measures or flagged by panelists themselves as being of general concern. These common themes were also classified according to relevant domain based on conversational context. This study was approved by the UCLA Institutional Review Board.

Recommendation of Measure

A measure was recommended by meeting nine minimum criteria based on DHCS-MHSD statutory requirements and

the main findings from each project stage. Criteria included broad age range (2–18 years); broad range of symptoms (internalizing and externalizing); availability in California's top three threshold languages (Spanish, Vietnamese, and Chinese); easy to use, as reported by the 56 county mental health agencies (mean score of ≥ 3 on a scale of 1, difficult, to 5, easy); brief in time of administration (< 10 minutes); consumer-centered version (parent or youth); acceptable strength of evidence (CEBM rating of ≤ 2); mean rating by Delphi panel of high or high-equivocal overall utility (≥ 6); and capacity to align time point to a unique episode of care (child's current treatment episode, which often varies by child).

RESULTS

Of the 49 state mental health agencies, 73% ($N=36$) reported use of at least one standardized screening measure, for an overall total of 15 unique measures (Table 1). Use of screening measures varied widely by age, with 11 states using a measure for children younger than 5, 13 states for children ages 5 to 18, and 18 states for young adults ages 19 to 21. Assessment spanned various domains, including development, symptoms, impairment, treatment goals, service intensity, and strengths. The most frequently reported screening measure was the Child and Adolescent Needs and Strengths (CANS) ($N=18$ states), followed by the Pediatric Symptom Checklist (PSC) ($N=9$), the Ages and Stages Questionnaire ($N=7$), and the Child and Adolescent Functional Assessment Scale (CAFAS) ($N=7$). Six states reported using their own custom-made measure. Of the 36 states that used at least one standardized screening measure, 10 (27%) reported use of at least one standardized measure to track clinical outcomes, for an overall total of six unique measures. Of these, only the CANS ($N=6$) and the CAFAS ($N=3$) were used by more than two states.

Among the 56 California county mental health agencies, the most frequently reported measures were the CANS ($N=33$), the Child Behavior Checklist (CBCL) ($N=14$), and the Eyberg Child Behavior Inventory (ECBI) ($N=12$) (Table 2). The reported purposes of the measures included screening, diagnosis, determining level of care, outcomes, treatment goals, and quality improvement. Most of the counties reported use of one measure for all these purposes, even if the purpose did not align with recommended use (e.g., using a service need intensity measure for diagnosis). In addition, 25 counties reported using tools that were not related to child functioning.

Of the 225 measures identified from the literature review, only 34 had been used in at least three published studies as a clinical outcome measure in a community-based mental health setting. Of these, seven measures remained after eliminating measures that were diagnosis specific ($N=20$), that were not applicable to the target population ($N=5$), or that did not measure change in clinical status over time ($N=2$) (see figure in online supplement). From other data sources, we identified four additional measures that were

TABLE 1. Standardized measures used in 36 states for assessment of mental health treatment among children, by purpose of assessment^a

Measure	Assessment or screening		Outcome	
	N	State	N	State
Child and Adolescent Needs and Strengths (CANS)	18	AL, GA, IN, LA, MA, ME, MD, MT, NH, NJ, NY, PA, TN, TX, VA, WA, WI, WV	6	IN, MA, MT, NH, PA, TX
Pediatric Symptom Checklist	9	IA, MA, MI, MN, ND, SC, TN, TX, VT	1	MN
Ages and Stages Questionnaire	7	CT, IA, MA, MI, ND, SC, UT		
Child and Adolescent Functional Assessment Scale (CAFAS)	7	HI, ID, KA, MI, NE, NM, NV	3	HI, ID, NV
Strengths and Difficulties Questionnaire	3	MA, MN, ND	0	
Child and Adolescent Service Intensity Instrument	3	AZ, HI, MN	2	HI, MN
Brief Infant-Toddler Social and Emotional Assessment	3	MA, MN ND	1	MN
Ohio Youth Problems, Functional, and Satisfaction scales ^b	2	IL, OH	1	IL
Youth Outcomes Questionnaire (YOQ)	2	AR, ME ^c	0	
Goal Attainment Scale	1	IL	0	
Behavioral and Emotional Rating Scale	1	OR	0	
Devereux Early Childhood Assessment Scale	1	IL	1	IL
Columbia Impairment Scale	1	IL	1	IL
Eyberg Child Behavior Inventory	1	TN	0	
Achenbach System of Empirically Based Assessment	1	VT	0	

^a States may report more than one measure. Sample excludes California. Six states (AK, CO, DE, FL, OH, and SD) exclusively used their own custom-designed tool; three used their own tool supplemented by a standardized tool (CAFAS, ID and KS; CANS, MD).

^b The Ohio scales were included because they are also used out of state.

^c Arkansas has discontinued use of the YOQ.

used by at least one state, were used by two or more California county mental health agencies, or were of interest to DHCS-MHSD. The final pool of 11 candidate measures included the Achenbach System of Empirically Based Assessment (ASEBA), which includes the CBCL; Clinical Global Impressions; the Strengths and Difficulties Questionnaire (SDQ); the CANS; the CAFAS; the ECBI; the PSC; the Treatment Outcome Package (TOP); the Children's Global Assessment Scale; the Ohio Youth Problems, Functional, and Satisfaction Scales (Ohio Scales); and the Youth Outcomes Questionnaire. (Details of how the candidate

measures met inclusion criteria are available in Supplemental Table 1 in the online supplement.)

With the CEBM protocol, the strength of evidence for the PSC, ASEBA, SDQ, and CAFAS was rated as a 2, corresponding to an individual cohort study. The other measures were rated as 4, corresponding to poor-quality cohort study, with the exception of the TOP, which had no outcome studies. (The psychometric properties and strength of evidence for use of each candidate measure as a clinical outcome measure in community-based mental health programs are summarized in online supplement Table 2.)

TABLE 2. County mental health agencies in California that reported use of a standardized measure for assessment of mental health treatment among children, by purpose of assessment^a

Measure	Any	Screening	Diagnosis	Level of care	Outcomes	Treatment goals	Quality improvement
Child and Adolescent Needs and Strengths	33	23	23	25	31	31	27
Child Behavior Checklist	14	11	11	11	12	11	11
Eyberg Child Behavior Inventory	12	10	9	8	12	10	10
Youth Outcomes Questionnaire	9	5	5	5	8	8	8
Child and Adolescent Level of Care Utilization System	7	5	4	7	6	4	5
Child and Adolescent Functional Assessment Scale	2	1	1	1	1	1	1
Pediatric Symptom Checklist	1	1	1	1	1	1	1
Other ^b	25	—	—	—	—	—	—

^a The sample represented 56 of California's 58 counties. Counties could report use of more than one measure.

^b Includes measures not related to functioning (i.e., child's early development) and disorder-specific measures.

TABLE 3. Ratings of 11 candidate measures by the modified Delphi panel (N=14)^a

Measure	Effective care			Scientific acceptability			Usability			Feasibility			Overall utility		
	M	SD	Range	M	SD	Range	M	SD	Range	M	SD	Range	M	SD	Range
Achenbach System of Empirically Based Assessment	7.7	.6	6–9	7.9	.8	5–9	6.5	1.4	4–8	6.6	1.0	3–8	7.3	.8	3–9
Strengths and Difficulties Questionnaire	6.5	1.5	5–7	6.2	1.5	3–7	6.9	1.8	4–8	7.3	1.9	4–9	6.6	1.7	4–8
Pediatric Symptom Checklist	7.1	1.2	4–9	7.5	1.3	6–8	7.2	1.2	5–8	7.3	1.2	5–9	6.3	1.5	2–8
Youth Outcome Questionnaire	5.9	.9	4–9	4.6	.8	4–7	5.2	1.8	3–9	5.1	1.9	5–8	4.9	1.2	4–9
Ohio Youth Problem, Functioning, and Satisfaction scales	4.9	1.2	4–7	3.9	1.2	2–7	4.1	1.3	4–7	4.7	1.4	4–7	4.3	1.0	4–7
Eyberg Child Behavior Inventory	4.6	1.8	4–8	5.6	1.5	4–8	5.2	2.0	4–8	5.1	2.0	3–8	3.9	1.4	4–6
Children's Global Assessment Scale	3.8	.8	2–8	3.7	.8	2–7	5.4	1.0	4–9	6.0	1.1	6–9	3.7	1.3	2–8
Child and Adolescent Needs and Strengths	4.4	.9	3–8	3.4	1.1	2–8	3.9	1.2	2–9	3.9	.9	2–8	3.5	1.3	2–9
Child and Adolescent Functional Assessment Scale	3.6	1.1	2–7	3.8	1.4	2–7	4.0	1.3	3–7	3.6	1.3	2–7	3.1	1.3	1–7
Clinical Global Impressions scale	2.9	1.5	1–6	2.6	1.2	2–7	3.9	1.6	2–6	4.6	1.5	3–9	2.6	1.6	1–6
Treatment Outcome Package	2.7	.9	1–7	2.1	1.0	1–6	3.0	1.3	2–8	2.6	1.5	1–7	2.3	1.2	1–7

^a Possible ratings range from 1 (lowest) to 9 (highest), with scores of 4 to 6 indicating equivocal ratings, except for overall utility, which is rated from 1 (definitely not recommend) to 9 (definitely recommend), with scores of 4 to 6 indicating equivocal utility.

Following deliberation and rerating by the modified Delphi panel, only the ASEBA, SDQ, and PSC were rated on average in the high-equivocal to high (≥ 6) range for use as a marker of effective care, scientific acceptability, usability, feasibility, and overall utility (Table 3). The remaining measures were rated consistently in the equivocal-to-low range, on average, for all domains. (Explanations for panel ratings and stakeholder priorities that emerged from the panel deliberation are summarized in online supplement Table 3.)

Panelists' priorities for a statewide performance measurement system included assessment of a broad range of symptoms, use with a wide age range, strong scientific evidence, availability in multiple languages, easy interpretation of findings, low burden to administer, parent report version, alignment with the current treatment episode, and timely feedback. The potential for the PSC to facilitate communication across primary care and specialty mental health care providers was viewed as a unique strength. Upon tallying the nine minimum criteria for recommendation for statewide use, only the PSC met all criteria (Table 4). Compared with the PSC, the ASEBA and SDQ, both of which met seven of the nine criteria, required longer time frames for evaluation (past 6 months for the ASEBA and past 6 months or current school year for the SDQ); as a result, they were well suited for detection of chronic symptoms but not for alignment with a child's unique episode of care.

DHCS-MHSD mandated the use of the PSC and CANS for the statewide performance measurement system (37). In fiscal year 2017–2018, \$14,952,000 was allocated to build a state-level data capture system and reimburse counties for the costs related to implementation of screening with CANS (i.e., training and clinician time to complete), information technology upgrades, and time spent preparing and submitting data to DHCS-MHSD. Implementation was phased in beginning July 1, 2018, starting with 32 counties and

followed by 26 additional counties beginning October 1, 2019, and by Los Angeles County beginning July 1, 2019.

DISCUSSION

The lack of a common approach for standardized outcomes measurement makes it impossible to compare child clinical outcomes across states and across counties within California. Only one out of five state mental health agency Web sites reported use of any standardized measure to track clinical outcomes for children receiving publicly funded mental health services, and only two reported any information on statewide implementation. At the state and California county level, the reported measures varied widely by child age, domains assessed, and format. California counties reported using standardized measures for a wide range of purposes, but none specified using a standardized clinical outcome measure to assess the effectiveness of care. In addition, the outcome measures reported at the state and county levels did not closely align with the strength of scientific evidence for use in community-based child mental health programs or with the Delphi panel ratings. The CANS was reported more frequently than any other measure at both the state and the county level, but the strength of its scientific evidence was poor and its Delphi panel ratings were low-equivocal to low across all domains. In contrast, the PSC, the second-most frequently reported measure among all states but used infrequently in California, rose to the top based on acceptable scientific evidence and high Delphi panel ratings and was the only measure that met all nine minimum criteria.

DHCS-MHSD's final selection struck a compromise by including the PSC because of its high rankings and the CANS because of its wide use in California. Implementation of the CANS was supported by funding to individual counties to cover the costs of clinician training and time. The CANS is

TABLE 4. Performance on nine criteria by 11 measures under consideration for assessment of mental health treatment among children in California

Measure	Broad age range (2–18)	Broad symptom range ^a	Top 3 threshold languages ^b	Easy to use	Brief (<10 minutes)	Consumer centered ^c	Acceptable strength of scientific evidence ^d	Overall utility ^e	Can align with current episode of care	Total ^f
Pediatric Symptom Checklist	+	+	+	+	+	+	+	+	+	9
Achenbach System of Empirically Based Assessment	+	+	+	+		+	+	+		7
Strengths and Difficulties Questionnaire	+	+	+		+	+	+	+		7
Child and Adolescent Functional Assessment Scale (CAFAS)	+	+	na	+	+		+		+	6
Youth Outcome Questionnaire		+	+	+	+	+				5
Eyberg Child Behavior Inventory				+	+	+			+	4
Child and Adolescent Needs and Strengths		+	+	+					+	4
Ohio Youth Problems, Functioning, and Satisfaction scales		+	+			+				3
Treatment Outcome Package		+			+	+				3
Clinical Global Impressions scale (CGI)			na		+					1
Children's Global Assessment Scale (CGAS)			na		+					1

^a Internalizing and externalizing behaviors.

^b Spanish (N=49 counties), Vietnamese (N=9), and any variety of Chinese (Cantonese, N=5; Mandarin, N=4; other Chinese, N=1). For tools with multiple informants, the criteria were satisfied if at least one version was available in the threshold language.

^c Parent or youth version.

^d The criterion was met if the measure was rated ≤2 on a 5-point scale, with 1 indicating individual randomized clinical trials with more than 80% follow-up and 5 indicating expert opinion or inconclusive evidence.

^e Mean Delphi panel rating of ≥6 out of 9, with 1 indicating would definitely not recommend, 4–6 indicating equivocal rating, and 9 indicating would definitely recommend.

^f Ratings for the CAFAS, CGI, and CGAS are underestimates because criteria were not considered to have been met if not applicable, but there was no change in major findings.

envisioned to facilitate communication by being a part of the clinical assessment (38) and is named in the legislation as an example of an “evidence-based model for performance outcome systems” (25). This approach is consistent with evidence that state legislators place higher priority on information from behavioral health organizations than from university-based research (39).

Successful implementation of a system for measurement of performance outcomes requires several components. Funds for a performance outcome measurement system were not earmarked in the state legislative mandate, which instead stipulates that DHCS minimize costs “by building upon existing resources to the fullest extent possible” (18), consistent with national trends (17). Meeting this mandate will require the development and maintenance of a relatively complex statewide data infrastructure that must include multiple clustering units of analyses (individual, provider, program, and agency), the documentation of use and fidelity

to evidence-based practices (including recommended medication treatment), approaches for case-mix adjustment and identification of disparities, and the capacity to link to clinical outcomes for children with variable episodes of care by using data sources that are not contingent upon continued contact with mental health services (5, 8, 12, 27, 40). Costs for this public investment in a measurement-driven system for assessing quality of care will be substantial and will require continual maintenance.

Other important considerations include specifying the purpose and corresponding unit of analysis (e.g., child, provider, program, county, state, or system) when selecting standardized measures to track clinical outcomes for children receiving publicly funded mental health services. The PSC was recommended because it satisfies the nine criteria identified as priorities for adopting a measure— it covers a broad age range; captures a wide breadth of symptoms; is available in California's top three threshold languages; is

easy to use; is brief and consumer centered; and has acceptable evidence strength, moderate to high overall utility, and a time period that can align with the child's unique episode of care.

Recommendation of the PSC should be viewed as complementary to other measurement-driven quality improvement activities (41) and does not preclude a program's use of other standardized measures that may be purposefully selected and individualized for routine outcome monitoring in clinical practice (42).

It is also important to choose a clinical outcome measure prior to developing a standardized set of methods and materials to electronically document and track the delivery of recommended care processes across county behavioral health agencies. Although the report to DHCS-MHSD provided some guidelines for implementation, development of the approach to collect and submit data were delegated to individual counties, potentially introducing greater heterogeneity in data quality when aggregated at the statewide level. Future research is needed to develop, maintain, and continuously refine statewide data infrastructure for monitoring the delivery of recommended care processes and their relationship to meaningful clinical outcomes as well as to track cost-shifting and potential savings across agencies serving children. As a starting point, it would be useful to develop a set of standardized materials and methods for data capture of the PSC and CANS by using a community-partnered approach in select counties. The data capture effort could then be pilot-tested and further refined prior to large scale use. For the PSC, such an approach could also capitalize on advances in digital health tools to enable primary caregivers and youths to report back in real time on clinical outcomes, ideally with results integrated into the electronic health care record (43). This would reduce selection bias because clinical outcomes monitoring would not be contingent on contact with mental health services.

CONCLUSIONS

A shared and consistent national mandate is required to provide equitable and effective care for children, while reducing costs and placing higher priority on child mental health care. The findings of this study illustrate the need for policy action to promote selection of a common clinical outcome measure and measurement methodology for children receiving publicly funded mental health care. Although this process will likely include advances and setbacks, a statewide performance outcome system remains an important component of systemwide goals.

AUTHOR AND ARTICLE INFORMATION

University of California, Los Angeles (UCLA), Semel Institute for Neuroscience and Human Behavior (Zima, Marti) and UCLA Center for Health Policy Research, Fielding School of Public Health (Lee, Pourat), UCLA. Send correspondence to Dr. Zima (bzima@mednet.ucla.edu). Research posters of preliminary findings were presented at the annual

meeting of the American Academy of Child and Adolescent Psychiatry, Washington, D.C., October 25, 2017; the World Psychiatric Association Thematic Congress, Melbourne, February 25–28, 2018; and the annual World Congress of Pediatrics, Fukuoka, Japan, July 30, 2018.

This study was funded by the California Health Care Foundation in support of the Department of Health Care Services (DHCS). The authors gratefully acknowledge the strong partnership with the DHCS mental health services division, the excellent work of the members of the modified Delphi panel, the comments from the subject matter expert panel during each stage of this project, and data verification support from Xiao Chen, Ph.D.

The authors report no financial relationships with commercial interests.

Received September 11, 2018; revision received November 30, 2018; accepted January 10, 2019; published online February 28, 2019.

REFERENCES

1. Patient Protection and Affordable Care Act. 42 USC §18001 et seq. 2010. <https://www.gpo.gov/fdsys/pkg/PLAW-111publ48/pdf/PLAW-111publ48.pdf>
2. About the National Quality Strategy. Rockville, MD, US Department of Health and Human Services, Agency for Healthcare Research and Quality, March 2017. <http://www.ahrq.gov/workingforquality/about.htm>
3. Berwick DM, Nolan TW, Whittington J: The triple aim: care, health, and cost. *Health Aff (Millwood)* 2008; 27:759–769
4. Burns BJ, Friedman RM: Examining the research base for child mental health services and policy. *J Ment Health Adm* 1990; 17: 87–98
5. Chassin MR, Loeb JM, Schmalz SP, et al: Accountability measures—using measurement to promote quality improvement. *N Engl J Med* 2010; 363:683–668
6. Children's Mental Health: An Overview and Key Considerations for Health System Stakeholders. Washington, DC, National Institute for Health Care Management, 2005
7. Gardner W, Kelleher KJ: Core quality and outcome measures for pediatric health. *JAMA Pediatr* 2017; 171:827–828
8. Glied SA, Stein BD, McGuire TG, et al: Measuring performance in psychiatry: a call to action. *Psychiatr Serv* 2015; 66:872–878
9. Institute of Medicine: Committee on Crossing the Quality Chasm: Adaptation to Mental Health and Addictive Disorders. Improving the Quality of Health Care for Mental and Substance-Use Conditions. Washington, DC, National Academy Press, 2006
10. Patel MM, Brown JD, Croake S, et al: The current state of behavioral health quality measures: where are the gaps? *Psychiatr Serv* 2015; 66:865–871
11. Pincus HA: Quality measures: necessary but not sufficient. *Psychiatr Serv* 2012; 63:523
12. Pincus HA, Spaeth-Rublee B, Watkins KE: The case for measuring quality in mental health and substance abuse care. *Health Aff (Millwood)* 2011; 30:730–736
13. Zima BT, Murphy JM, Scholle SH, et al: National quality measures for child mental health care: background, progress, and next steps. *Pediatrics* 2013; 131(suppl 1):S38–S49
14. Seibert J, Fields S, Fullerton CA, et al: Use of quality measures for Medicaid behavioral health services by state agencies: implications for health care reform. *Psychiatr Serv* 2015; 66:585–591
15. Children's Health Care Quality Measures. Baltimore, Medicaid.gov. <https://www.medicare.gov/medicaid/quality-of-care/performance-measurement/child-core-set/index.html>. Accessed April 5, 2018
16. Blackburn J, Becker DJ, Morrissey MA, et al: An assessment of the CHIP/Medicaid quality measure for ADHD. *Am J Manag Care* 2017; 23:e1–e9
17. Pincus HA, Scholle SH, Spaeth-Rublee B, et al: Quality measures for mental health and substance use: gaps, opportunities, and challenges. *Health Aff (Millwood)* 2016; 35:1000–1008

18. Realignment Revisited: An Evaluation of the 1991 Experiment in State-County Relations. Sacramento, CA, Legislative Analyst Office, 2001. http://www.lao.ca.gov/2001/realignment/020601_realignment.html
19. Mental Health Realignment. Sacramento, CA, Legislative Analyst Office, 2011. http://www.lao.ca.gov/handouts/Health/2011/Mental_Health_1_26_11.pdf
20. Arnquist S, Harbage P: A Complex Case: Public Mental Health Delivery and Financing in California. Oakland, California Health Care Foundation, 2013. <http://www.chcf.org/~media/MEDIA%20LIBRARY%20Files/PDF/PDF%20C/PDF%20ComplexCaseMentalHealth.pdf>
21. Rosenblatt A, Rosenblatt J: Demographic, clinical, and functional characteristics of youth enrolled in six California systems of care. *J Child Fam Stud* 2000; 9:51–66
22. Rosenblatt A, Attkisson CC: Integrating systems of care in California for youth with severe emotional disturbance: I. a descriptive overview of the California AB377 evaluation project. *J Child Fam Stud* 1992; 1:93–113
23. Rosenblatt A, Rosenblatt JA: Assessing the effectiveness of care for youth with severe emotional disturbances: is there agreement between popular outcome measures? *J Behav Health Serv Res* 2002; 29:259–273
24. Rosenblatt A, Attkisson CC: Integrating systems of care in California for youth with severe emotional disturbance, III. answers that lead to questions about out-of-home placements and the AB377 evaluation project. *J Child Fam Stud* 1993; 2:119–141
25. Performance Outcomes System Statute: Welfare and Institutions Code, Section 14707.5. <https://codes.findlaw.com/ca/welfare-and-institutions-code/wic-sect-14707-5.html>. Accessed Feb 7, 2019
26. Meisel J: Mental Health Services Oversight and Accountability Commission (MHSOAC) Evaluation Master Plan. Sacramento, CA, MHSOAC, 2013. http://archive.mhsoc.ca.gov/Evaluations/docs/EvaluationMasterPlan_Final_040413.pdf
27. Ashwood JS, Kataoka SH, Eberhart NK, et al: Evaluation of the Mental Health Services Act in Los Angeles County: Implementation and Outcomes for Key Programs. Santa Monica, CA, RAND Corp, 2018. https://www.rand.org/pubs/research_reports/RR2327.html
28. Promises Still to Keep: A Decade of the Mental Health Services Act. Sacramento, CA, Little Hoover Commission, 2015. <https://mentalillnesspolicy.org/wp-content/uploads/LittleHooverCommish.pdf>
29. Performance Outcomes System Plan for Medi-Cal Specialty Mental Health Services for Children and Youth. Sacramento, CA, Department of Health Care Services, 2015. http://www.dhcs.ca.gov/individuals/Documents/POS_LegReport_05_15.pdf
30. Pourat N, Zima B, Marti A, et al: California Child Mental Health Performance Outcomes System: Recommendation Report. Los Angeles, UCLA Center for Health Policy Research, 2017. <http://healthpolicy.ucla.edu/publications/search/pages/detail.aspx?PubID=1660>
31. Brook RH: The RAND/UCLA Appropriateness Method; in *Methodology Perspectives*. Edited by McCormick K, Moore S, Siegel R. AHCPR pub no 95-0009. Rockville, MD, Agency for Health Care Policy & Research, 1994
32. Delphi Method. Santa Monica, CA, RAND Corp. <https://www.rand.org/topics/delphi-method.html>. Accessed April 21, 2016
33. National Quality Forum: Tool Evaluation Criteria and Guidance Summary Tables Effect for Projects Beginning After January 2011. www.qualityforum.org/docs/Measure_Evaluation_Criteria_Guidance.aspx. Accessed April 2, 2016
34. Levels of Evidence. Oxford, United Kingdom, Centre for Evidence-based Medicine, 2009. <https://www.cebm.net/2009/06/oxford-centre-evidence-based-medicine-levels-evidence-march-2009/>
35. Ochs E: Transcription as theory. *Developmental Pragmatics* 1979; 10:43–72
36. Ryan GW, Bernard HR: Techniques to identify themes. *Field Methods* 2003; 15:85–109
37. Medi-Cal Specialty Mental Health Services. Sacramento, California Department of Health Care Services, 2018. http://www.dhcs.ca.gov/services/Pages/Medi-cal_SMHS.aspx
38. Anderson RL, Lyons JS, Giles DM, et al: Reliability of the Child and Adolescent Needs and Strengths–Mental Health (CANS-MH) scale. *J Child Fam Stud* 2003; 12:279–289
39. Purtle J, Dodson EA, Nelson K, et al: Legislators' sources of behavioral health research and preferences for dissemination: variations by political party. *Psychiatr Serv* 2018; 69:1105–1108
40. Kilbourne AM, Keyser D, Pincus HA: Challenges and opportunities in measuring the quality of mental health care. *Can J Psychiatry* 2010; 55:549–557
41. Fortney JC, Unützer J, Wrenn G, et al: A tipping point for measurement-based care. *Psychiatr Serv* 2017; 68:179–188
42. Boswell JF, Kraus DR, Miller SD, et al: Implementing routine outcome monitoring in clinical practice: benefits, challenges, and solutions. *Psychother Res* 2015; 25:6–19
43. Archangeli C, Marti FA, Wobga-Pasiah EA, et al: Mobile health interventions for psychiatric conditions in children: a scoping review. *Child Adolesc Psychiatr Clin N Am* 2017; 26:13–31