

Article

Prioritizing quality measure concepts at the interface of behavioral and physical healthcare

HAROLD ALAN PINCUS^{1,2}, MINGJIE LI³, DEBORAH M. SCHARF⁴,
BRIGITTA SPAETH-RUBLEE³, MATTHEW L. GOLDMAN^{3,5},
PARASHAR P. RAMANUJ⁶, and ERIN K. FERENCHICK⁷

¹Department of Psychiatry, College of Physicians and Surgeons, Columbia University, 1051 Riverside Drive, Unit 09, New York, NY 10032, USA, ²New York-Presbyterian Hospital, 630 West 168th Street, New York, NY 10032, USA, ³New York State Psychiatric Institute, 1051 Riverside Drive, Unit 9, New York, NY 10032, USA, ⁴Department of Psychology, Lakehead University, 955 Oliver Road, Thunder Bay Ontario, P7B 5E1, Canada, ⁵Department of Psychiatry, Columbia University Medical Center, New York State Psychiatric Institute, 1051 Riverside Drive, Box 99, New York, NY 10032, USA, ⁶Royal National Orthopaedic Hospital, 45 Bolsover Street, London, W1W 5AQ, UK, and ⁷Center for Family and Community Medicine, Columbia University Medical Center, 610 West 158th Street, New York, NY 10032, USA

Address reprint requests to: Harold Alan Pincus, Department of Psychiatry, College of Physicians and Surgeons, Columbia University, 1051 Riverside Drive, Unit 09, New York, NY 10032, USA. Tel: +1-646-774-5400; Fax: +1-646-774-8443; E-mail: pincush@nyspi.columbia.edu

Editorial Decision 25 May 2017; Accepted 31 May 2017

Abstract

Objective: Integrated healthcare models can increase access to care, improve healthcare quality, and reduce cost for individuals with behavioral and general medical healthcare needs, yet there are few instruments for measuring the quality of integrated care. In this study, we identified and prioritized concepts that can represent the quality of integrated behavioral health and general medical care.

Design: We conducted a literature review to identify candidate measure concepts. Experts then participated in a modified Delphi process to prioritize the concepts for development into specific quality measures.

Setting: United States.

Participants: Expert behavioral health and general medical clinicians, decision-makers (policy, regulatory and administrative professionals) and patient advocates.

Main outcome measures: Panelists rated measure concepts on importance, validity and feasibility.

Results: The literature review identified 734 measures of behavioral or general medical care, which were then distilled into 43 measure concepts. Thirty-three measure concepts (including a segmentation strategy) reached a predetermined consensus threshold of importance, while 11 concepts did not. Two measure concepts were 'ready for further development' ('General medical screening and follow-up in behavioral health settings' and 'Mental health screening at general medical healthcare settings'). Among the 31 additional measure concepts that were rated as important, 7 were rated as valid (but not feasible), while the remaining 24 concepts were rated as neither valid nor feasible.

Conclusions: This study identified quality measure concepts that capture important aspects of integrated care. Researchers can use the prioritization process described in this study to guide healthcare quality measures development work.

Key words: mental health, quality measurement, behavioral health care, general medical health care, integrated healthcare

Introduction

Behavioral health (i.e. mental health and substance use conditions) and general medical services are often fragmented. The fragmentation leads to unmet healthcare needs, poor health outcomes and high societal costs [1]. For instance, in the United States, mental illness is nearly universal among the highest cost, most frequently hospitalized Medicaid beneficiaries [2]. Similarly, among the nine countries surveyed by the Commonwealth Fund, patients with three or more chronic diseases and functional disability use more healthcare services than other groups and often experience healthcare coordination problems [3].

To resolve these issues, many countries are integrating general medical and behavioral health services. In England, integrated primary and acute care systems are emerging that join general practice, hospital, community and mental health services [4]. In Germany, sickness funds offer integrated care contracts and disease management programs for chronic illnesses to improve coordination among providers in the ambulatory sector and to improve patient care [5]. In Canada, initiatives such as the Strategy for Patient-Oriented Research network integrates primary prevention and healthcare services horizontally and vertically across the care continuum [6]. In the United States, approaches such as the collaborative care model have shown cost savings in low-income and high-risk populations relative to usual care [7].

Large-scale uptake of these models, however, has been slow for several reasons, including a lack of incentives for significant system overhaul, and a lack of accompanying payment infrastructure and improvement strategies to sustain quality care after significant changes to the system have been made [8]. In particular, there are few behavioral, and even fewer integrated behavioral and general medical, healthcare quality measures available to support integrated healthcare systems reform. Of the 646 measures endorsed by the National Quality Forum (NQF), only 53 (8%) are behavioral health measures; of these, 14 (2%) reflect integrated care [9]. Similarly, among the 34 quality measures for Medicare and Medicaid Services accountable care organizations, only 2 (6%) are behavioral health measures [10]. To date, few studies define or operationalize appropriate care, and service data and existing health information technology (IT) are often insufficient to support behavioral health or integrated care reporting [11].

To help create the infrastructure needed to promote the uptake of integrated healthcare models, we sought to initiate the development of quality measures at the interface of general medical and behavioral health. Specifically, we used a modified Delphi process to prioritize integrated care measure concepts spanning all three components of Donabedian's model of healthcare quality (structure, process and outcomes) for future use in value-based purchasing, public reporting and quality improvement in general [12].

Methods

Overview

The Delphi method, developed by the RAND Corporation, is an iterative process that uses statistical presentations of survey data to systematically create expert consensus [13]. The Delphi method has been used effectively in other studies of quality measure development [14, 15]. In this study, we used a modified, two-round Delphi process to create expert consensus on the importance, validity and feasibility of integrated care quality measure concepts.

Measure and measure concepts

In the initial phase of this study, the authors conducted a comprehensive literature review to identify existing measures in behavioral health and primary care [16]. We searched keywords related to physical and behavioral care to identify relevant measures from the NQF (www.qualityforum.org/) and National Quality Measures Clearinghouse (www.qualitymeasures.ahrq.gov/) up to September 2015. In total, we identified 734 measures, most of which pertained to primary care (not behavioral health or integrated care).

Next, the authors distilled the 734 measures into 43 'measure concepts,' i.e. concepts that could be the foundation for future fully specified measures. Measure concepts are not fully specified and tested; instead, they represent a group of indicators that measure similar aspects of healthcare [17]. For example, to transform the concept 'general medical ongoing assessment' (Box 2) into a set of operationalized measures will require measure developers to specify which 'general medical conditions' should be included and to define the type of 'ongoing assessment' needed for each condition. Measure concepts in our study apply to general medical or behavioral health settings (or both).

We also asked panelists to evaluate a quality measure 'segmentation strategy.' Specifically, panelists rated how much they agreed that existing or endorsed quality metrics should be calculated separately for patients with behavioral health conditions (i.e. named within the denominator). This approach can inform direct comparisons between the quality of care for the full patient population and that for patients with behavioral comorbidities [18]. Research has shown that people with behavioral health conditions (especially those with more severe conditions) often have limited access to both general medical and behavioral healthcare, and that they experience poorer quality of care, resulting in worse health outcomes and significantly decreased lifespan. This 'segmentation strategy' could help ensure that quality care was being provided to this historically under-served group.

Survey instrument

Panelists were asked to rate 43 measure concepts that were organized into 3 sections and 12 domains (Box 1). The three sections, defined by healthcare settings and corresponding patient populations, included: (i) individuals with serious behavioral health conditions (i.e. serious mental illness and/or substance use disorder) and at risk for general medical conditions, seen in behavioral health settings; (ii) individuals with mild-to-moderate behavioral health conditions, seen in general medical settings and (iii) cross-cutting concepts applicable at integrated behavioral and general medical settings. The survey was administered via Qualtrics online (cumc.co1.qualtrics.com) and is available upon request. An example of the survey format is in Box 2.

The rating criteria for each measure concept ('importance,' 'validity' and 'feasibility') were adapted from measure evaluation criteria applied by the NQF [19]. A concept was considered important if it addressed a clear gap between the actual and potential level of care, applied to many individuals, and/or there were serious adverse consequences from not adhering to the concept. A concept was considered valid if adequate scientific evidence or professional consensus existed to suggest that the measure-specified performance of care is linked to patient health. A concept was considered feasible if the information necessary to implement the measure was available in administrative data, the medical record or patient survey and the additional costs of capturing the data would not place a large burden on the clinical setting. Panelists rated importance, validity and feasibility

Box 1 Delphi survey questionnaire outline

General medical care in behavioral health settings

- Domain 1: General medical screening or diagnostic assessment and prevention
- Domain 2: Outcomes and patient perception of care
- Domain 3: Continuity and coordination of care
- Domain 4: Access to general medical care

Behavioral healthcare in general medical settings

- Domain 5: Behavioral health screening or assessment
- Domain 6: Behavioral health evidence-based treatment
- Domain 7: Behavioral health patient-centered care
- Domain 8: Continuity and coordination of care
- Domain 9: Access to behavioral healthcare

Cross-cutting measure concepts at the interface of behavioral health and general medical settings

- Domain 10: Continuity and coordination of care
- Domain 11: Social service access
- Domain 12: Cost and efficiency

Segmentation strategy: Stratification of general medical intervention by behavioral health comorbidity

Box 2 Delphi survey questionnaire item format—example of measure concept rated by Delphi panel

Measure concept: general medical ongoing assessment

- Numerator: Individuals receiving specified recommended ongoing assessment for a chronic general medical condition (e.g. blood pressure monitoring for hypertension)
- Denominator: Individuals in the behavioral health disparities subpopulation with the relevant chronic general medical condition that requires follow-up

Importance	1	2	3	4	5	6	7	8	9
Validity	1	2	3	4	5	6	7	8	9
Feasibility	1	2	3	4	5	6	7	8	9

Other/Comments:

on a 9-point Likert scale, with scores of 1–3 indicating strong disagreement, 4–6 as uncertain agreement and 7–9 as strong agreement.

Delphi process

Having an appropriate composition of panelists in consensus methodology is vital for determining the legitimacy of the results [20]. We chose panelists with expertise in behavioral health and/or general medical quality assessment, and who were (i) leaders in their clinical/policy fields as evidenced by their roles within national organizations (e.g. federal agencies, State Medicaid programs and health service research organizations) or (ii) peers recommended by leaders in the field. An initial list of 31 potential members was created by the research team, among which 25 panelists ultimately took part in at least one round of survey. Among the 23 panelists who participated in the final round of survey, 6 were behavioral healthcare clinicians, 9 were general medical care clinicians, 13 were decision-makers (individuals working in policy, regulatory, or administrative settings) and 1 was patient advocacy group representative. (Some

panelists were counted more than once if they belonged to more than one panel category.)

Panelists twice rated each measure concept for importance, validity and feasibility. In the first round, panelists reported their ratings through a survey. Feedback and results of the first rating round were circulated to all panelists, who then met for a face-to-face meeting to reflect on the issues driving their ratings and to clarify reasons for disagreements. The meeting maintained a specific structure: a moderator set the context and engaged all panelists in an interactive and balanced dialog to contain the influence of dominant personalities. After discussion, panelists re-rated the initial list of measure concepts, taking the discussion into account. Those who were absent from the meeting received the meeting notes and audio record to inform their second-round surveys.

Analysis

We first computed Cronbach’s α to measure the consistency with which panelists rated the importance, validity and feasibility of

Table 1 Prioritization framework to categorize integrated care quality measure concepts based on ratings for each criterion

Category	Importance	Validity	Feasibility	Implication	Strategy
A	High ^a	High	High	Ready for measure development	Operationalize and test specific measures based on the concept
B	High	High	Low ^b	Information not available in existing data and hard to capture	Potential measurement innovation; build the data system capacity
C	High	Low	High/low	No existing evidence to support the link between performance and measure Evidence demonstrated lack of association between measure and outcomes	Conduct research to explore the association between measure and outcomes Exclude for further development (unless related to balancing measure development)
D	Low	High/low	High/low	Did not meet criteria for importance and had limited support in other dimensions	Exclude for further development (unless related to balancing measure development)

^aHigh is defined as reached consensus on that criterion (at least 75% of Delphi panel members rated 7 or higher on a 9-point scale).

^bLow is defined as failed to reach consensus on that criterion (less than 75% of Delphi panel members rated 7 or higher on a 9-point scale).

Table 2 Percentage of integrated care quality measure concepts reached consensus on each criterion

Rating criterion	Delphi round 1 (<i>n</i> = 24)			Delphi round 2 (<i>n</i> = 23)		
	Consensus percentage	Total mean	Cronbach's α	Consensus percentage	Total mean	Cronbach's α
Importance	57	7.3	0.94	75	7.5	0.95
Validity	16	6.5	0.94	20	6.7	0.94
Feasibility	9	6.1	0.93	5	6.0	0.92

measure concepts across the two survey rounds [21]. We then computed the mean value and standard deviation of each rating criterion (importance, validity or feasibility) for each measure concept. Measure concepts were deemed to have reached consensus on a rating criterion if at least 75% panel members rated a measure 7 or higher on the 9-point scale [22]. Based on the consensus status of each rating criterion, we then categorized each measure concept's readiness to be developed into specific quality measures by using a prioritization framework (Table 1). Category A concepts achieved consensus on importance, validity and feasibility, and are therefore ready for further development. Category B concepts reached consensus on importance and validity, but not feasibility (e.g. due to gaps in health IT impeding data collection). Category C concepts met consensus on importance only but not on validity or feasibility (e.g. data do not show or does not exist to show an association between service and health). Concepts were placed in Category D if they did not achieve consensus on importance, regardless of their ratings for validity or feasibility.

Results

In total, 24 panelists participated in the first-round survey and 23 took the second-round survey (77% and 72% response rates, respectively). Proportionally, more measure concepts achieved consensus on importance while less than a quarter of the measure concepts achieved consensus on validity or feasibility during each survey round. From the first to the second round of ratings, the percentage of measure concepts reaching consensus on importance increased by 18%. Internal consistency was high on each one of the three criteria in both rounds (Cronbach's α greater than 0.9 indicating an excellent degree of consistency) [23] (Table 2).

Final results were analyzed based on the data from the second round of survey. Of all 44 survey items (including 43 measure concepts and the segmentation strategy), 33 achieved the predetermined threshold for consensus on importance. Out of those 33 items, the top 5 (ranked by importance mean value) were in the 'Behavioral health screening or assessment in general medical settings' domain.

Overall, only two measure concepts reached consensus on all three criteria. The segmentation strategy had a high mean score on importance and reached consensus on both importance and validity. The measure concepts are sorted into prioritization categories according to the prioritization framework (Table 3).

Measure concepts applicable to behavioral healthcare settings

Among measure concepts applicable to behavioral health settings, only 'General medical screening and follow-up' met the predetermined threshold for consensus on all three criteria (Category A). Both 'General medical ongoing assessment' and 'General medical outcomes' were in Category B. Six measure concepts were in Category C and seven measure concepts in Category D (which the panel did not deem important).

Measure concepts applicable to general medical care settings

Among measure concepts applicable to general medical settings, only 'Mental health screening' was rated suitable for further development (Category A). Four concepts were in Category B: 'Mental health follow-up,' 'Substance use disorder assessment,' 'Suicide risk follow-up' and 'Mental health systematic ongoing assessment.' Twelve concepts failed to meet the threshold for consensus on validity and feasibility (Category C) and Category D included three measure concepts.

Cross-cutting measure concepts applicable at integrated behavioral and general medical settings

No measure concepts in this section were in Category A or B. Six cross-cutting measure concepts were in Category C and one concept in Category D. None were rated sufficiently valid or feasible to meet the threshold suggesting the need for research testing the link between integrated care procedures and health outcomes.

Table 3 Final rating mean scores and prioritization categories for integrated care quality measure concepts

Measure concept	Importance Mean (SD)	Validity Mean (SD)	Feasibility Mean (SD)	Category ^a
GM care in BH settings				
GM screening and follow-up	7.7 (1.37)	7.5 (1.44)	7.3 (1.29)	A
Scheduled GM ambulatory/primary care visit within 12 months	6.7 (2.30)	6.4 (1.83)	7.0 (1.80)	D
GM ongoing assessment	8.1 (0.90)	7.8 (1.04)	7.1 (1.44)	B
GM outcomes	8.2 (0.95)	7.6 (0.95)	6.7 (1.19)	B
Patient harm episodes	6.8 (1.40)	5.9 (1.68)	5.3 (2.00)	D
Inappropriate medications	7.5 (1.04)	6.7 (1.18)	5.4 (1.62)	C
Patient/Family experiences of GM care	7.6 (1.67)	6.6 (0.99)	5.9 (1.22)	C
Provider-patient communication around GM Care	7.9 (1.08)	6.1 (1.69)	5.0 (1.60)	C
Patient mortality	7.7 (1.51)	6.6 (1.53)	5.9 (1.66)	D
BH inpatient discharge and follow-up for GM comorbidity	7.4 (1.15)	6.4 (0.99)	5.5 (1.47)	C
Penetration/Utilization of care (GM care received)	6.8 (1.70)	6.3 (1.54)	6.2 (1.48)	D
Duration of untreated GM condition	6.4 (2.17)	5.4 (1.95)	3.7 (1.64)	D
Patient experience of access to GM services	7.7 (0.86)	6.7 (1.06)	5.7 (1.33)	C
Tracking referrals for GM/surgical care in BH settings	7.8 (1.03)	6.9 (1.25)	5.4 (1.16)	C
Duration of BH hospitalization	5.9 (1.50)	5.8 (1.51)	7.0 (1.40)	D
Readmission for BH hospitalizations	7.3 (1.29)	6.7 (1.15)	6.9 (1.53)	D
BH care in GM settings				
Mental health screening	8.3 (0.69)	7.9 (0.85)	7.3 (1.02)	A
Mental health follow-up	8.4 (0.65)	7.4 (1.27)	5.8 (1.04)	B
SUD assessment	8.3 (0.86)	7.6 (1.12)	6.9 (0.97)	B
SUD follow-up	8.3 (0.82)	7.1 (1.32)	5.6 (1.16)	C
Suicide risk screening	7.9 (0.95)	6.9 (0.79)	6.5 (0.90)	C
Suicide risk follow-up	8.4 (0.66)	7.5 (0.90)	6.0 (1.49)	B
Mental health systematic ongoing assessment/monitoring	8.1 (0.87)	7.4 (0.90)	6.6 (1.04)	B
BH medication persistence	7.5 (0.95)	6.7 (1.02)	6.7 (1.49)	C
Stepped care	7.3 (1.14)	6.2 (1.77)	5.0 (1.66)	C
Psychotherapy utilization threshold	6.9 (1.35)	6.0 (1.60)	5.4 (1.62)	D
Provider-patient communication around BH care	7.8 (1.34)	6.0 (1.68)	4.9 (1.58)	C
Patient/Family experiences of BH care	8.0 (0.80)	6.9 (1.18)	6.0 (1.46)	C
GM inpatient discharge and follow-up for BH comorbidity	7.6 (1.12)	6.4 (1.27)	5.8 (1.40)	C
Penetration/Utilization of care (BH care received)	6.4 (1.95)	6.1 (1.83)	6.5 (1.75)	D
Patient experience of access to BH services	7.7 (1.01)	6.7 (1.11)	6.0 (1.41)	C
Duration of untreated BH condition	7.0 (1.69)	6.0 (1.74)	4.7 (1.72)	C
Behavioral health wait times	7.7 (1.10)	6.6 (1.27)	5.4 (1.38)	C
Tracking referrals for BH specialist care in GM settings	8.0 (0.82)	7.2 (1.11)	5.4 (1.27)	C
Duration of GM hospitalization	5.9 (1.55)	5.7 (1.48)	7.1 (1.41)	D
Readmissions for GM hospitalizations	7.5 (1.16)	6.9 (1.12)	7.2 (1.41)	C
Cross-cutting measure concepts at the interface of BH and GM settings				
Coordination between GM and BH Care	7.8 (0.89)	6.3 (1.74)	4.7 (1.61)	C
Medication reconciliation	8.2 (0.90)	6.5 (1.65)	6.4 (1.27)	C
Assessment of social services	8.0 (0.88)	7.1 (1.20)	5.8 (1.23)	C
Receipt of social services	7.8 (1.19)	6.7 (1.43)	5.1 (1.08)	C
Average cost of care per patient	7.6 (1.27)	6.9 (1.46)	6.8 (1.68)	C
Out-of-Pocket costs	7.3 (1.21)	6.2 (1.30)	5.0 (1.35)	C
Use of low-value interventions	6.1 (1.81)	5.5 (1.53)	5.3 (1.45)	D
Segmentation strategy				
Stratification of GM intervention by BH comorbidity	8.0 (0.67)	7.6 (0.99)	6.6 (1.59)	B

GM, general medical; BH, behavioral health; SUD, substance use disorder.

^aBased on the prioritization framework (Table 1). Category A concepts reached consensus (at least 75% panelists rated 7 or higher) on importance, validity and feasibility; Category B concepts reached consensus on importance and validity; Category C reached consensus on importance; Category D did not reach consensus on importance.

Discussion

Implications for future measure development

Experts agreed that most of the measure concepts included in this study were important. In contrast, however, only a small minority of measure concepts was deemed valid or feasible. At the Delphi panel meeting, some panelists explained the reasons. A number of

panelists expressed the difficulty in rating feasibility of measure concepts without fully specified measures, as some parameters could greatly affect their use (e.g. 'Duration of untreated behavioral health conditions in general medical care settings'). They also reported providing low feasibility ratings because collecting relevant data would be burdensome or because behavioral health services were not well-

specified in administrative data [24]. The low ratings of validity, in some cases, can be attributed to evidence of a weak association between the measure concept and health outcomes or the lack of evidence to support a significant association. Different concerns regarding the same measure concept resulted in moderate rating on validity and low feasibility or vice versa. For example, on the measure concept ‘Patient experience of access to general medical services in behavioral health settings,’ one expert commented that ‘Patient experience is accurate on this but [the likelihood of] low response rates could limit feasibility’; while another expert addressed that ‘Patients may not know what is supposed to be happening, and thus the validity is questionable.’ Also, many panelists expressed concern about the possible association between validity and feasibility, as discussed by other investigators using the Delphi method [25].

These overall ratings show that the proposed set of measure concepts appropriately reflects the most urgent integrated care quality measurement needs. The obvious lack of consensus on validity and feasibility suggests areas for future empirical research and infrastructure development, respectively. Categorizing the proposed measure concepts according to the prioritization framework provides clear guidance on next steps, be it toward measure development and pilot testing, further research to strengthen the evidence base of a measure concept or building data infrastructure, or moving attention away from measures that were considered unimportant by the panel.

The two measure concepts that were deemed ready for further development (‘General medical screening and follow-up’ in behavioral health settings and ‘Mental health screening’ in general medical settings) reflect the implementation of integrated care already underway. That is, experts saw the importance, validity and feasibility of screening, and in some cases monitoring, of conditions typically treated in other settings. However, experts were also unlikely to believe that aspects of care beyond screening and monitoring are valid or feasible measures of the quality of integrated care; in other words, they questioned whether this monitoring and reporting will lead to appropriate actions in the current healthcare system. For instance, evaluations of the first 56 grant recipients for the Substance and Mental Health Services Administration’s Primary and Behavioral Healthcare Integration grants project showed that behavioral health settings successfully integrate general medical screening and monitoring into their workflow, but had less success tracking and improving access to general medical care not directly provided by the agency’s own staff [26, 27]. This alignment between actual practice and expert opinion lends validity to the results presented here, with experts acknowledging that systems and providers would already be reporting on Category B (important and valid but not feasible) processes if they could. It also suggests the need for agencies to build the data system technical and structural capacity to routinely capture behavioral healthcare quality data.

Category D measure concepts were rated as unimportant. Although the prioritization framework suggests excluding them from further research or development, some of them could be used to develop ‘balancing measures,’ which can provide additional information to facilitate the interpretation of other quality measures. For example, a potential balancing measure concept was ‘Penetration/Utilization of care’ (in behavioral healthcare received by patients in general medical setting, or vice versa), which can be measured to clarify how to interpret other quality measures in populations with varying rates of utilization of care.

Panelists also rated the value of assessing healthcare quality separately for persons with behavioral health conditions (‘segmentation strategy’) given known disparities in access to and receipt of quality care. Delphi panelists rated the segmentation strategy as important and valid but not feasible, likely due to challenges associated with

defining the disparity population. As such, researchers should work to clearly define this disparity group. Once this is underway, they should consider developing companions to existing NQF-endorsed measures with a denominator specifically defined for individuals with behavioral health conditions. Policymakers might then include this new ‘disparities category’ in the National Healthcare Quality and Disparities Reports.

Limitations

Although the authors carefully selected panelists to reflect a broad range of expertise, their views do not necessarily completely represent their corresponding stakeholder groups. There was also disproportionate representation from some stakeholder groups, with only one representative from a patient advocacy group. The proposed measure concepts cover broad themes for quality measurement rather than specific and scientifically tested numerator and denominator definitions, leaving this work to measure developers. Finally, two of the panelists helped develop the survey; however, sensitivity analyses showed that including their ratings had no significant impact on how any measure concept was categorized.

Conclusion

Integrated care models are emerging as a promising strategy to improve access and provide better healthcare to patients with complex behavioral health and general medical needs, as well as reducing their healthcare costs. Large-scale uptake of these models requires a set of robust quality indicators sensitive to resource demands of an integrated healthcare system. We used a modified Delphi process to prioritize quality measure concepts related to several important aspects of integrated care. We also developed and applied a framework that delineates next steps for developing measures of integrated care. Although this study was conducted in the context of the United States, measure concepts can be adapted to other health systems. The methodology used for measure concept development and the prioritization framework could also be applied more broadly, beyond health or behavioral health integration.

Acknowledgements

The authors would like to thank the panelists who took part in the Delphi study: Adam Wilcox (Medical Informatics, Intermountain Healthcare); Emma Beth McGinty (Department of Health Policy and Management, Johns Hopkins University); Carolyn Clancy (Health for Organizational Excellence, U.S. Department of Veteran Affairs); David Baker (Division of Health Care Quality Evaluation, The Joint Commission); Erin Ferenchick (Co-author of this article; Center for Family and Community Medicine, Columbia University); Frank deGruy (Department of Family Medicine, University of Colorado Denver); Gail Daumit (Department of Medicine, Division of General Internal Medicine, Johns Hopkins University School of Medicine); Helen Burstin (National Quality Forum); Henry Chung (Montefiore Care Management Organization); Jeffrey Buck (Center for Clinical Standards and Quality, Centers for Medicare & Medicaid Services); Joe Parks (Missouri HealthNet; National Council for Behavioral Health); Judy Zerzan (Colorado Department of Health Care Policy and Financing); Kevin Larsen (Centers for Medicare & Medicaid Services); Kirsten Beronio (The Center for Medicaid and CHIP Services, Centers for Medicare & Medicaid Services); Laura Jacobus-Kantor (Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration); Lauren Murray (Consumer Engagement and Community Outreach, National Partnership for Women & Families); Lisa Kern (Department of Medicine, Department of Healthcare Policy and Research, Weill Cornell Medicine); Lisa

Patton (Division of Evaluation, Analysis and Quality, Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration); Matthew Goldman (Co-author of this article; Columbia University, New York State Psychiatric Institute); Nirav Shah (Kaiser Permanente); Rachel Talley (Columbia University, New York State Psychiatric Institute); Sapana Patel (Columbia University, New York State Psychiatric Institute); Sarah Scholle (Research & Analysis, National Committee for Quality Assurance); Timothy Ferris (Population Health Management, Partners HealthCare); Virna Little (Psychosocial Services and Community Affairs, The Institute for Family Health).

Funding

This publication was supported by the Commonwealth Fund (grant number 20141104). The views presented here are those of the authors and not necessarily those of The Commonwealth Fund or its directors, officers or staff. Additional funding was provided by the National Center for Advancing Translational Sciences, National Institutes of Health, through Grant Number UL1 TR000040. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

References

- Horvitz-Lennon M, Kilbourne AM, Pincus HA. From silos to bridges: meeting the general health care needs of adults with severe mental illnesses. *Health Aff* 2006;25:659–69.
- Boyd C, Leff B, Weiss C *et al*. Clarifying multimorbidity patterns to improve targeting and delivery of clinical services for Medicaid populations. *Center Health Care Strat* 2010. http://www.chcs.org/media/Clarifying_Multimorbidity_for_Medicaid_report-FINAL.pdf (1 January 2017, date last accessed).
- Sarnak DO, Ryan J. How high-need patients experience the health care system in nine countries. *Commonwealth Fund* 2016;1:1–14.
- National Health Service in England. New care models: Vanguard – developing a blueprint for the future of NHS and care services 2016. Available from: https://www.england.nhs.uk/wp-content/uploads/2015/11/new_care_models.pdf.
- The Commonwealth Fund. What is being done to promote delivery system integration and care coordination? [Available from: <http://international.commonwealthfund.org/features/integration/>].
- Canadian Institutes of Health Research. SPOR Network in Primary and Integrated Health Care Innovations: Background, goals, objectives, governance and priority focus. 2014.
- Unützer JHH, Schoenbaum M, Druss B. The collaborative care model: an approach for integrating physical and mental health care in Medicaid health homes. *Center Health Care Strat* 2013. http://www.chcs.org/media/HH_IRC_Collaborative_Care_Model_052113_2.pdf (1 January 2017, date last accessed).
- Goldman ML, Spaeth-Rublee B, Pincus HA. Quality indicators for physical and behavioral health care integration. *JAMA* 2015;314:769–70.
- National Quality Forum. Measures, reports, and tools. 2016. Available from: http://www.qualityforum.org/measures_reports_tools.aspx.
- Centers for Medicare & Medicaid Services. Accountable Care Organization 2016 Program Quality Measure Narrative Specifications. 2016.
- Kilbourne AM, Keyser D, Pincus HA. Challenges and opportunities in measuring the quality of mental health care. *Can J Psychiatry* 2010;55:549–57.
- Donabedian A. The quality of care: how can it be assessed? *JAMA* 1988;260:1743–8.
- Fink A, Koseoff J, Chassin M *et al*. Consensus methods: characteristics and guidelines for use. *Am J Public Health* 1984;74:979–83.
- Kröger E, Tourigny A, Morin D *et al*. Selecting process quality indicators for the integrated care of vulnerable older adults affected by cognitive impairment or dementia. *BMC Health Serv Res* 2007;7:195.
- Minkman MM, Ahaus KT, Huijsman R. A four phase development model for integrated care services in the Netherlands. *BMC Health Serv Res* 2009;9:42.
- Goldman ML, Spaeth-Rublee B, Nowels AD *et al*. Quality measures at the interface of behavioral health and primary care. *Curr Psychiatry Rep* 2016;18:39.
- National Quality Forum. Multiple Chronic Conditions Measurement Framework. 2012. http://www.qualityforum.org/Publications/2012/05/MCC_Measurement_Framework_Final_Report.aspx (6 June 2017, date last accessed).
- Pence BW, O'Donnell JK, Gaynes BN. The depression treatment cascade in primary care: a public health perspective. *Curr Psychiatry Rep* 2012;14:328–35.
- National Quality Forum. Measure evaluation criteria [Available from: http://www.qualityforum.org/Measuring_Performance/Submitting_Standards/Measure_Evaluation_Criteria.aspx].
- Campbell SM, Cantrill JA. Consensus methods in prescribing research. *J Clin Pharm Ther* 2001;26:5–14.
- Bland JM, Altman DG. Statistics notes: Cronbach's alpha. *BMJ* 1997;314:572.
- Boulkedid R, Abdoul H, Loustau M *et al*. Using and reporting the Delphi method for selecting healthcare quality indicators: a systematic review. *PLoS One* 2011;6:e20476.
- Gliem JA, Gliem RR, editors. Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales. 2003: Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education. <http://www.ssnstudents.com/wp/wp-content/uploads/2015/02/Gliem-Gliem.pdf> (1 January 2017, date last accessed).
- Ranallo PA, Kilbourne AM, Whatley AS *et al*. Behavioral health information technology: from chaos to clarity. *Health Aff* 2016;35:1106–13.
- Peña A, Virk SS, Shewchuk RM *et al*. Validity versus feasibility for quality of care indicators: expert panel results from the MI-Plus study. *Int J Qual Health Care* 2010;22:201–9.
- Scharf DM, Eberhart NK, Schmidt Hackbarth N, *et al*. Evaluation of the SAMHSA Primary and Behavioral Health Care Integration (PBHCI) Grant Program. 2014. http://www.rand.org/content/dam/rand/pubs/research_reports/RR500/RR546/RAND_RR546.pdf (1 January 2017, date last accessed).
- Scharf DM, Schmidt Hackbarth N, Eberhart NK *et al*. General medical outcomes from the Primary and Behavioral Health Care Integration grant program. *Psychiatr Serv* 2016;67:1226–32.