ORIGINAL RESEARCH

Degree of Behavioral Health Integration and Patient Outcomes

Benjamin Littenberg, MD, Levi Bonnell, MPH, PhD, Peter Callas, PhD, Juvena Hitt, MPH, C. R. Macchi, PhD, Matt Martin, PhD, MS, LMFT, CSSBB, Mindy L. McEntee, PhD, Daniel J. Mullin, PsyD, MPH, Gail L. Rose, PhD, and Constance van Eeghen, DrPH, MHSA, MBA

Background: **Primary care practices with greater integration of behavioral health care have better patient-reported outcomes. We sought to identify whether there is a threshold effect in the relationship between the degree of Integrated Behavioral Health (IBH) and patient-reported outcomes.**

Methods: **Secondary analysis of survey results from Integrating Behavioral Health and Primary Care, a multistate longitudinal randomized, controlled study of 3,929 adults with multiple chronic medical and behavioral conditions. Patient outcomes included Patient-Reported Outcomes Measurement Information System-29 (PROMIS-29) functional status (PROMIS-29), depression (PHQ-9), anxiety (GAD-7), the Duke Activity Status Index, Consultation and Relational Empathy (CARE), patient centeredness, and utilization. IBH was measured by the Practice Integration Profile (PIP) version 1.0. The optimal threshold was identified by examining the relationship of PIP to PROMIS-29. The discriminatory power of the threshold was examined using multilevel linear regression with adjustment for potential confounders.**

Results: **Fifteen of 44 practices with 1,237 patients were highly integrated (PIP $\geq$ 65). All outcomes tended to be better in patients from practices with high integration. After adjustment for potential confounders, the relationship remained beneficial for all outcomes, with Pain Intensity (−0.51 [95% CI −0.97, 0.04]), patient centeredness (2.52 [0.88, 4.16]), and CARE (1.62 [0.62, 2.61]) statistically significant.**

Conclusions: **Patients in high integration practices report better outcomes. A measurable target for IBH, such as a PIP total score $\geq$ 65, provides a focus for practice leadership and guidance on the time and resources needed to achieve integration associated with positive patient outcomes. The results of this analysis provide further evidence of the broad, beneficial impacts of integrating behavioral health and primary care services. (J Am Board Fam Med 2025;38:834–845.)**

Keywords: **Behavioral Medicine, Integrated Delivery of Health Care, Patient Reported Outcomes, Practice Management, Primary Health Care, Psychometrics, Surveys and Questionnaires**

Introduction

Behavioral Health (BH) care encompasses mental health, substance use, health behavior, and support for psychological and social issues. Although primary care offers a prime opportunity to address these needs[1], a lack of resources and systems to deliver these services often leaves patients without needed care[2], despite efficacious interventions.[3–9]

This article was externally peer reviewed.

Submitted 8 February 2025; revised 23 April 2025; accepted 5 May 2025.

From the University of Vermont, Burlington, VT (BL, PC, JH, GLR, CE); Society of Thoracic Surgeons, Chicago, IL (LB); Arizona State University, Phoenix, AZ (MM, MLM); UMass Chan Medical School, Worcester, MA (DJM).

Funding: This work was funded through a Patient-Centered Outcomes Research Institute (PCORI) Award (PCS-1409-24372). The views, statements, and opinions presented in this report are solely the responsibility of the authors and do not necessarily represent the views of PCORI, its Board of Governors or Methodology Committee. PCORI is an independent, nonprofit organization authorized by Congress in 2010. Its mission is to fund research that will provide patients, their caregivers, and clinicians with the evidence-based information needed to make better-informed healthcare decisions. PCORI is committed to continually seeking input from a broad range of stakeholders to guide its work.

Conflict of interest: The authors have no competing or conflicting interests to declare.

Corresponding author: Benjamin Littenberg, MD, Robert Larner, M.D. College of Medicine, 89 Beaumont Ave., Burlington, VT 05405 (E-mail: benjamin.littenberg@uvm.edu).

An arbitrary separation of physical and behavioral health has perpetuated siloed care and is inconsistent with the lived experience of most patients, particularly those with multiple chronic conditions.[10]

Despite evidence supporting Integrated Behavioral Health (IBH), just 39% of family physicians report collaborative work with behavioral health clinicians (BHCs).[11] Implementation varies, with different approaches to integration in different practices[6,12] and wide variability in the degree of integration.[13] Considerable unevenness in definitions, access, conditions addressed, care delivered, and the extent of integration into practice workflow make analysis of IBH efforts difficult.[11,14]

Peek's Lexicon[15] established broadly accepted core dimensions of integration, defining IBH as "A practice team of primary care and behavioral health clinicians working together with patients and families, using a systematic and cost-effective approach, to provide patient-centered care for a defined population." Common models of integration such as the Collaborative Care Model or the Primary Care Behavioral Health model vary in their emphasis on different elements of this definition. Real-world integration frequently combines selected features of multiple models, ranging from simple colocation (BHC has the same address as medical clinicians) to intensive integration of practice workflow, work space, infrastructure, records, support systems, consults, and professional education.[16]

The Practice Integration Profile (PIP), derived from the 2013 version of the Lexicon, is the first model-agnostic measure of integration to be psychometrically validated.[13,17,18] There is a positive and statistically significant association between degree of integration (measured by PIP total score) and patient function. For example, the PROMIS-29 Mental Health Summary was 0.05 points higher for every point of Total PIP ($P = .05$).[19] This relationship has implications for clinical, organizational, and financial operations as well as policy. Although more integration is associated with better outcomes, we do not know how much integration is needed. We sought to identify whether there is a threshold effect in the relationship between the degree of IBH and patient-reported outcomes in primary care, that is, a level of integration beyond which outcomes are significantly better and whether such a threshold predicts future patient outcomes.

Methods

Data Sources

Secondary data were analyzed from Integrating Behavioral Health and Primary Care (IBH-PC), a large study of adult primary care patients with multiple chronic conditions from 2016 to 2021.[19–21] Data were collected from practice staff, clinicians and patients from each of 44 family medicine and internal medicine practices with colocated BHCs in 13 US states at baseline and follow-up 2 years later. Patients had at least 3 medical conditions or 1 behavioral and 1 medical condition from the following list: arthritis, obstructive lung disease, diabetes, heart disease (heart failure or hypertension), mood disorder (anxiety or depression), chronic pain (headache, migraine, neuralgia, fibromyalgia, or chronic musculoskeletal pain), insomnia, irritable bowel syndrome, and substance misuse (substance use disorder, tobacco use, or problem drinking). Conditions were identified via electronic medical record using International Classification of Disease diagnosis codes, medications, problem lists, and laboratory results. Participants were included without regard to whether they had received behavioral health services, which was not recorded. Participants self-reported their demographics, functional status, and other outcomes via online, mail, or telephone surveys in English or Spanish.

Instruments

The Patient-Reported Outcomes Measurement Information System-29 (PROMIS-29)[22] is a 29-item questionnaire with 8 domains: physical function, anxiety, depression, fatigue, sleep disturbance, social functioning, pain intensity, and pain interference. Lower scores indicate better outcomes for depression, anxiety, fatigue, pain interference, pain intensity, and sleep disturbance. Higher scores indicate better physical function and social participation. Physical and Mental Health Summaries are calculated that combine the 8 domains and rescale them all so that higher scores represents better function.[23] Scores are standardized to the US adult population, with mean of 50 and standard deviation of 10. The minimally important difference for scales in the PROMIS series is 2.0 to 5.0.[24,25]

The Patient Health Questionnaire-9 (PHQ-9)[26] and Generalized Anxiety Disorder-7 (GAD-7)[27] measure depression and anxiety symptom severity on continuous scales from 0 to 27 and 0 to 21. Higher scores indicate higher symptom burden.

Patient-centered primary care was assessed by the Patient Centeredness Index[28], a 14-item survey that records how patients perceive their care. The Consultation and Relational Empathy (CARE) scale is a 10-item survey used to assess patients' perceptions of clinician empathy.[29] Both instruments report higher scores for practices with better performance.

The Duke Activity Status Index[30] is a 12-item self-reported measure of functional capacity that correlates well with maximal oxygen consumption. We converted the index to Metabolic Equivalent of Task (METs) units[31]; higher METs indicate better function.

The Utilization Patient Report[32] is a 3-item survey assessing health care utilization in the past year, including emergency department visits, overnights in the hospital, and outpatient appointments. The Restricted Activity Days[33] survey assesses restriction of daily life due to illness and disability.

PIP 1.0 is a 30-item survey of clinicians, nurses, administrators, and staff to assess the degree to which BH and primary care services are integrated within a practice.[13,34] Four or 5 nurses, staff, and behavioral and medical clinicians (at least 1 of each) from each site completed the PIP. Most of the questions start with "In our practice. . .", followed by a specific characteristic (such as ". . .we use registry tracking for patients with identified BH issues"), an example ("Insomnia registry"), a definition ("Numerator=# of patients in BH registries; Denominator=# of patients with BH needs"), and 5 responses from Never to Always. It has 6 domains: workflow, clinical services, workspace, clinician integration, patient identification, and patient engagement, with 2 to 9 questions each. Scores range from 0 (least degree of integration) to 100. The Total Integration Score represents the overall degree of behavioral health integration.

In previous analyses using data not otherwise reported here, the PIP 1.0 has a Cronbach's α of 0.95 with high retest reliability.[13,35] It was tested for validity and reliability in a sample of 1,372 respondents from 774 practices in 52 states/territories. The total PIP score was 58 with a standard deviation of 23, median of 61, and range from 0 to 100. Within each practice, the median range in total score among respondents was 10.[13] An exploratory factor analysis found an α coefficient >0.80 for 5 of the domains. We used the median value of the responses from each practice. Other practice characteristics were reported by clinic management or derived from census data for the practice's county.

Analysis

Identifying a Threshold

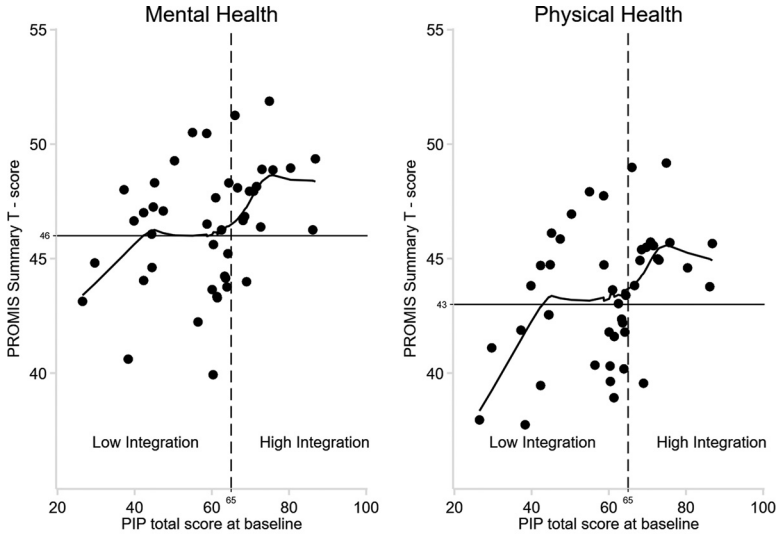
We sought to identify a Total PIP score threshold that would identify practices with good clinical outcomes with high specificity. In other words, above the threshold, most practices should have better-than-average clinical outcomes, and few practices should have poor outcomes. Baseline PIP was plotted against baseline mean PROMIS-29 mental and physical health summary scores for each practice and the mean PROMIS-29 summary score was plotted as a horizontal line. We overlaid locally-weighted smoothing curves (LOWESS)[36] to enhance visualization. Based on visual inspection, we selected a vertical threshold at PIP = 65 that isolated a sizable number of practices above the horizontal with few below (Figure 1). Similarly, a receiver-operating characteristic (ROC) curve plotting the true-positive rate (sensitivity) of the PIP for better practices against the false-positive rate (1-specificity) demonstrated good discrimination for both the Mental and Physical Health Summary Scores near a PIP of 65.

Modeling the Effect of the Threshold

Unadjusted multilevel linear regression models were used to estimate the mean differences in each patient-reported outcome in high- versus low-integration practices with 95% confidence intervals. Practice was included as a random intercept to account for similarities of patients within practices. Potential confounders included the number of qualifying diagnoses; age; gender; race coded as white versus Nonwhite; ethnicity coded as Hispanic versus non-Hispanic; marital status coded as married or living as married versus widowed, separated, divorced or single; education coded as college attendance versus high school or less; income <$30,000 per year; employment coded as full-time, part-time, student or homemaker versus retired, disabled, or unemployed. Each was added to the model as a fixed effect if they changed the coefficient of high- versus low-integration practices on the outcome by $\pm 10\%$[37,38] in unadjusted models and was associated with both the predictor and the outcome with $P < .15$.

The threshold for statistical significance was set at 5%. Analyses were performed in Stata 18.0 (StataCorp LP, College Station, TX, USA). Institutional Review Board approval was obtained before data collection.

Figure 1. Relationship of baseline Patient-Reported Outcomes Measurement Information System-29 (PROMIS-29) summary scores and Practice Integration Profile (PIP) scores.



Notes: Each dot represents the mean (at the practice level) of the PROMIS-29 Mental (left panel) or Physical (right) Health Summary T-score and the median of the total PIP score. Higher scores represent better patient function. The vertical dashed lines represent the threshold. The horizontal dashed lines represent the mean patient reported score across all 44 practices. The solid lines portray the smoothed (LOWESS) relationship between PIP and PROMIS-29.[36]

Results

Both the PROMIS-29 Mental Health and Physical Health Summary scores tended to increase with greater integration. (Figure 1) Visual inspection showed that both plots had a point near a Total PIP score of 65 that best distinguished low from high integration practices. Fifteen of 44 participating primary care practices (34%) scored 65 or better at baseline (high-integration practices) and cared for 31% of the patients. 93% of the high-integration practices had average Mental Health Summary scores above 46 and Physical Health Summary scores above 43, indicating that a PIP > 65 had high specificity for practices with good outcomes. High- and low-integration practices were similar in all recorded practice-level characteristics, except for PIP scores. (Table 1). However, patients in high-integration practices were significantly older, male, white, non-Hispanic, married, employed, and had more education, higher incomes, and fewer chronic medical and behavioral problems (Table 2).

Practices that had total PIP scores above 65 at baseline had better unadjusted patient outcomes at baseline (Table 3). For instance, among the more integrated sites, the PROMIS-29 Physical Health Summary was 2.7 points higher at baseline ($P = .001$) and the Mental Health Summary was 2.5 points higher ($P = .001$). Beneficial effects were seen for all PROMIS-29 domains, clinician empathy, functional capacity, depression, anxiety, functional capacity, emergency department visits, and patient centeredness. Health care utilization and restricted activity days showed better outcomes in the high-integration practices, but did not reach statistical significance. After adjustment for patient and neighborhood characteristics, all the effects remained beneficial, but were smaller. Pain Intensity, patient-centeredness, and empathy remained statistically significant. (Figure 2)

A similar pattern was observed at follow-up 2 years later. Unadjusted differences favored the high-integration practices in every outcome. After adjustment for patient and neighborhood characteristics, the differences still favored high integration, but were smaller; social participation, physical function, the Physical Health Summary, patient centeredness, empathy, and the Duke Activity Status Index were statistically significant (Table 3).

Table 1. Practice Characteristics

	All Practices	Low-Integration Practices (PIP < 65)	High-Integration Practices (PIP ≤ 65)	
	Mean (±Standard Deviation) or n (%)			*P*-value*
Practice characteristics				
N	44	29	15	
Visits per year	27,298 (±19,178)	25,052 (±15,098)	31,639 (±25,359)	0.48
Primary care clinician count	9.9 (±5.6)	10.2 (±6.0)	9.1 (±5.1)	0.73
Baseline primary care clinician full-time equivalents	6.0 (±3.1)	6.7 (±3.5)	4.8 (±1.8)	0.09
Baseline behavioral health clinician full-time equivalents	1.5 (±1.1)	1.5 (±1.2)	1.5 (±0.8)	0.89
Panel size	9319 (±4969)	9344 (±4677)	9271 (±5664)	0.87
% Adult medicare	20.8 (±14.1)	18.4 (±12.2)	25.6 (±16.7)	0.15
Tenure of on-site behavioral health services, y	6.8 (±6.9)	7.0 (±7.2)	6.5 (±6.7)	0.77
County social deprivation index	45.2 (±21.6)	46.2 (±22.1)	43.2 (±21.2)	0.45
Median age of county	37.3 (±3.5)	37.4 (±3.8)	37.1 (±2.8)	0.87
County proportion male	48.8% (±1.6%)	48.9% (±1.5%)	48.7% (±1.8%)	0.91
County proportion white	72.9% (±18.0%)	75.3% (±16.8%)	68.4% (±19.9%)	0.29
County proportion black	9.4% (±11.4%)	8.5% (±10.4%)	10.9% (±13.4%)	0.56
County proportion Asian	6.1% (±8.4%)	5.9% (±9.5%)	6.4% (±6.0%)	0.12
County proportion American Indian/ Alaskan Native/Pacific Islander	0.8% (±1.2%)	0.7% (±0.6%)	1.1% (±1.9%)	0.87
County proportion other race	5.9% (±5.7%)	5.4% (±4.8%)	6.8% (±7.1%)	0.49
County proportion Hispanic	15.6% (±13.8%)	14.8% (±13.5%)	16.9% (±15.0%)	0.37
High school graduation rate of county	89.2% (±4.1%)	89.5% (±2.6%)	88.5% (±6.1%)	0.61
County median income	64,516 (±15,612)	63,002 (±16,815)	67,443 (±13,007)	0.17
Population per square mile	1571 (±4907)	929 (±940)	2812 (±8348)	0.81
Practice specialty				0.63
Internal medicine	7 (16%)	4 (14%)	3 (20%)	
Family medicine	22 (50%)	16 (55%)	6 (40%)	
Mixed	15 (34%)	9 (31%)	6 (40%)	
Community health center	15 (34%)	11 (38%)	4 (27%)	0.52
Hospital-based	22 (50%)	13 (45%)	9 (60%)	0.53
Privately owned	4 (9%)	2 (7%)	2 (13%)	0.60
Academic site	21 (48%)	14 (48%)	7 (47%)	>0.99
Training site	18 (41%)	12 (41%)	6 (40%)	>0.99
Non-profit clinic	39 (89%)	26 (90%)	13 (87%)	>0.99
Geographic region				0.43
Pacific Northwest	3 (7%)	1 (3%)	2 (13%)	
Mountain	8 (18%)	7 (24%)	1 (7%)	
The South	8 (18%)	5 (17%)	3 (20%)	
New England	10 (23%)	5 (17%)	5 (33%)	
Mid Atlantic & Great Lakes	6 (14%)	5 (17%)	1 (7%)	
West Coast & Hawaii	9 (21%)	6 (21%)	3 (20%)	
Urban county	37 (84%)	24 (83%)	13 (87%)	>0.99
Total practice integration profile score at baseline	59.3 (±14.2)	52 (±11.2)	73.3 (±6.5)	<0.001
Range	27 to 87	27 to 64	66 to 87	

Abbreviation: PIP, Practice Integration Profile.
*Unadjusted comparisons by Wilcoxon rank-sum or χ^2 tests.

Table 2. Patient Characteristics

	All Practices	Low-Integration Practices (PIP < 65)	High-Integration Practices (PIP ≤ 65)	
	Mean (±standard deviation) or n (%)			*P*-value*
N	3,929	2,692	1,237	
Age at baseline†	61.9 (±13.2)	61.6 (±13.2)	62.6 (±13.4)	0.01
Women†	63.1%	64.5%	60.0%	0.007
Race†				<0.001
White	75.3%	73.7%	78.8%	
Black or African-American	11.8%	13.8%	7.6%	
American Indian or Alaska native	1.0%	0.9%	1.2%	
Asian	3.5%	3.2%	4.2%	
Native Hawaiian/other Pacific Islander	1.4%	1.1%	1.9%	
Other	7.0%	7.3%	6.2%	
Ethnicity†				0.008
Non-Hispanic	91.1%	90.3%	92.8%	
Hispanic	7.9%	8.8%	5.9%	
Prefer not to say	1.1%	1.0%	1.2%	
Marital status†				<0.001
Never married	15.8%	16.5%	14.3%	
Married	45.1%	41.9%	51.9%	
Living as married	2.6%	2.6%	2.5%	
Separated	2.7%	3.1%	1.9%	
Divorced	21.7%	23.1%	18.8%	
Widowed	12.0%	12.7%	10.6%	
Employment†				<0.001
Full-time	20.0%	18.1%	23.9%	
Part-time	8.6%	8.2%	9.4%	
Retired	37.4%	36.0%	40.3%	
Disabled	26.3%	29.4%	19.8%	
Homemaker	3.6%	3.8%	3.1%	
Student	0.9%	0.9%	0.9%	
Unemployed/Looking	3.2%	3.5%	2.4%	
Other	0.1%	0.1%	0.2%	
Highest level of education†				<0.001
<9th grade	3.1%	3.6%	2.1%	
9th–12th grade	8.6%	10.4%	4.5%	
High school diploma or equivalent	41.6%	42.2%	40.1%	
Associate's degree	15.9%	15.9%	15.8%	
Bachelor's degree	16.2%	14.0%	20.8%	
Graduate degree	14.7%	13.7%	16.7%	
Annual income <$30,000†	51.0%	57.1%	38.4%	<0.001
Diagnoses‡				
Hypertension	80.0%	81.2%	77.3%	0.004
Irritable bowel syndrome	4.2%	4.7%	3.2%	0.03
Insomnia	22.5%	22.9%	21.7%	0.42
Anxiety	31.9%	32.2%	31.4%	0.61
Arthritis	38.9%	42.1%	32.0%	<0.001
Asthma	20.4%	22.3%	16.1%	<0.001
Chronic obstructive lung disease	13.5%	16.1%	7.7%	<0.001
Chronic Pain	78.3%	81.3%	71.9%	<0.001
Diabetes	44.9%	46.7%	40.9%	<0.001

Continued

Table 2. Continued

	All Practices	Low-Integration Practices (PIP < 65)	High-Integration Practices (PIP ≥ 65)	
	Mean (±standard deviation) or n (%)			*P*-value*
Heart failure	8.0%	9.3%	5.3%	<0.001
Depression	44.2%	46.6%	38.9%	<0.001
Substance use disorder	22.0%	25.8%	13.8%	<0.001
Tobacco use	17.6%	21.4%	9.2%	<0.001
Alcohol use disorder	6.3%	7.2%	4.3%	<0.001
Number of qualifying conditions†	4.2 (±1.8)	4.4 (±1.8)	3.6 (±1.6)	<0.001

Abbreviation: PIP, Practice Integration Profile.
*Unadjusted comparisons by Wilcoxon rank-sum or χ^2 tests.
†Included as possible confounding effect.
‡Diagnostic groups are not mutually exclusive.

Discussion

In adjusted analyses, patients in practices that have achieved a Total PIP score of at least 65 at baseline reported better functional and other outcomes for nearly all domains examined, with significant differences in pain intensity, patient centeredness, and empathy. Results were similar at follow-up 2 years later, with significant effects seen for social participation, physical function, Physical Health Summary, patient centeredness, empathy, and the Duke Activity Status Index. Because the adjusted models showed smaller effects than the unadjusted regressions, it seems likely that some, but not all, of the effects may be attributable to social and demographic differences.

Anywhere in the moderate range of PIP, from about 45 to 65, is associated with a similar level of functional status as measured by the PROMIS (Figure 1). The mechanism of this phenomenon is uncertain, but may suggest that Behavioral Health services require a "critical mass" of integration to be effective.

Better outcomes in more integrated practices were observed at the practice level even though not all the patients in each practice received direct care from a behavioral health clinician. Patients in practices with higher levels of behavioral health integration reported greater levels of empathy from their PCPs and a greater sense that their care was patient-centered. These aspects of the therapeutic relationship benefit patient outcomes.[39,40] The close collaboration between medical and behavioral clinicians required for IBH may positively influence the ability of the team to provide patient-centered care[41] even when the BHC is not directly involved. To be patient-centered, a practice must function well as a team and communicate with patients in a way that empowers the patient to ask questions and participate in treatment decisions.[42,43] Integrated care involves the entire practice team of clinicians and staff working together with patients and families to provide patient-centered care.[44] The association between integration and patient centeredness may be based on the interconnected definitions of these 2 constructs.

Although the relationship between integrated care and patient health, cost of care, and improved experience of care has been demonstrated,[7,8,45] the present study is unique: the level of integration of each practice was assessed using a validated measure; it focused on the practice-level effect of integration, including patients who did and did not have contact with a BHC; and it addressed both physical and mental health outcomes. Prior research often focused on models serving narrow, specific populations, like the Collaborative Care Model.[46] Our study, in contrast, takes a pragmatic approach to delivering multiple interventions to diverse populations within a practice, with broad impacts on practice culture, care delivery, and meaningful patient centered outcomes across a broad range of subjects.

Measurement of integration on a continuous scale, as provided by the Total PIP score, is useful not just for research but for planning, designing, implementing, and monitoring integration. Goal setting with objective measures and feedback is critical for effective management.[47] A valid, specific target – a Total PIP score of 65, for example – can support the implementation of IBH by health care leaders, managers, clinicians, and staff using consistent measurement to effect complex change.[48–50]

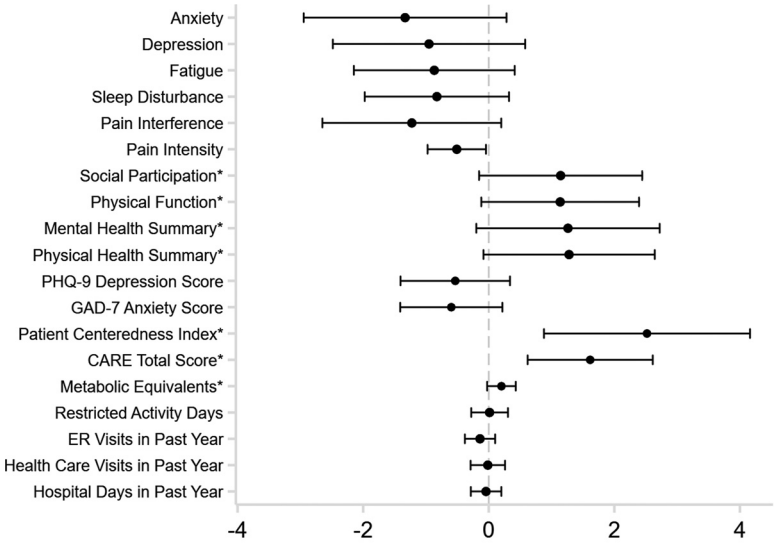
Table 3. Effect of High Baseline Integration on Patient-Reported Outcomes

	Unadjusted				Adjusted			
Outcome	N	Effect	95% Confidence Interval	*P*	N	Effect	95% Confidence Interval	*P*
At baseline								
PROMIS-29 functional status								
Anxiety	3,908	−2.28	−3.84, −0.72	0.004	3,705	−1.33	−2.95, 0.28	0.11
Depression	3,911	−1.93	−3.41, −0.44	0.01	3,708	−0.95	−2.48, 0.58	0.22
Fatigue	3,906	−1.87	−3.05, −0.68	0.002	3,704	−0.87	−2.15, 0.41	0.18
Sleep disturbance	3,908	−1.66	−2.82, −0.51	0.005	3,706	−0.83	−1.97, 0.32	0.16
Pain interference	3,906	−2.47	−3.90, −1.04	0.001	3,703	−1.22	−2.65, 0.20	0.09
Pain intensity	3,902	−0.85	−1.34, −0.35	0.001	3,699	−0.51	−0.97, −0.04	0.032
Social participation*	3,908	2.39	1.09, 3.70	<0.001	3,706	1.15	−0.15, 2.44	0.08
Physical function*	3,911	2.45	0.99, 3.91	0.001	3,708	1.14	−0.12, 2.39	0.08
Physical health summary*	3,893	2.67	1.12, 4.22	0.001	3,691	1.28	−0.08, 2.64	0.07
Mental health summary*	3,893	2.46	1.03, 3.90	0.001	3,691	1.26	−0.20, 2.72	0.09
PHQ-9 depression score	3,790	−1.27	−2.18, −0.35	0.007	3,622	−0.53	−1.40, 0.34	0.23
GAD-7 anxiety score	3,824	−1.10	−1.95, −0.26	0.01	3,659	−0.60	−1.41, 0.22	0.15
Patient centeredness index*	3,638	3.02	1.41, 4.62	<0.001	3,594	2.52	0.88, 4.16	0.003
Consultation and relational empathy total score*	3,891	0.18	0.08, 0.28	<0.001	3,710	1.62	0.62, 2.61	0.001
Restricted activity days	3,808	−0.14	−0.42, 0.13	0.32	3,641	0.01	−0.28, 0.31	0.93
Duke activity status index, (METs)*	3,616	0.48	0.20, 0.75	0.001	3,489	0.2	−0.03, 0.43	0.08
Utilization								
Emergency room visits in past year	3,873	−0.29	−0.54, −0.04	0.02	3,692	−0.14	−0.38, 0.10	0.26
Health care visits in past month	3,869	−0.19	−0.48, 0.11	0.21	3,698	−0.01	−0.29, 0.26	0.92
Hospital days in past year	3,864	−0.18	−0.41, 0.05	0.12	3,695	−0.04	−0.29, 0.20	0.73
At Follow-up								
PROMIS-29 functional status								
Anxiety	3,194	−1.94	−3.16, −0.71	0.002	3,150	−1.18	−2.44, 0.08	0.07
Depression	3,194	−1.79	−3.11, −0.46	0.008	3,150	−1.01	−2.37, 0.35	0.14
Fatigue	3,193	−1.88	−3.07, −0.69	0.002	3,062	−0.84	−2.21, 0.53	0.23
Sleep disturbance	3,194	−1.57	−2.67, −0.46	0.005	3,150	−0.69	−1.89, 0.51	0.26
Pain interference	3,193	−2.19	−3.48, −0.89	0.001	3,062	−0.88	−2.19, 0.44	0.19
Pain intensity	3,196	−0.72	−1.18, −0.27	0.002	3,064	−0.39	−0.80, 0.02	0.07
Social participation*	3,191	2.70	1.44, 3.96	<0.001	3,061	1.51	0.20, 2.83	0.024
Physical function*	3,195	2.82	1.51, 4.14	<0.001	3,064	1.46	0.21, 2.72	0.022
Physical health summary*	3,189	2.87	1.54, 4.19	<0.001	3,059	1.53	0.19, 2.87	0.025
Mental health summary*	3,189	2.22	0.99, 3.45	<0.001	3,059	1.26	−0.15, 2.68	0.08
PHQ-9 depression score	3,085	−1.15	−1.88, −0.42	0.002	2,960	−0.6	−1.36, 0.15	0.12
GAD-7 anxiety score	3,107	−0.79	−1.44, −0.15	0.02	3,063	−0.42	−1.06, 0.21	0.19
Patient centeredness index*	3,137	2.02	0.30, 3.75	0.02	3,093	1.78	0.04, 3.53	0.045
Consultation and relational empathy total score*	3,154	0.13	0.02, 0.23	0.02	3,118	1.16	0.13, 2.19	0.028
Restricted activity days	3,091	−0.15	−0.36, 0.06	0.17	2,977	−0.06	−0.28, 0.16	0.58
Duke activity status index, (METs)*	2,991	0.53	0.26, 0.80	<0.001	2,869	0.26	0.02, 0.49	0.033
Utilization								
Emergency room visits in past year	3,143	−0.23	−0.41, −0.05	0.01	3,016	−0.14	−0.32, 0.04	0.12
Health care visits in past month	3,139	−0.25	−0.55, 0.04	0.10	3,019	−0.16	−0.46, 0.13	0.28
Hospital days in past year	3,144	−0.16	−0.40, 0.08	0.19	3,024	−0.08	−0.31, 0.15	0.51

*Higher scores represent desired outcomes. *P*-values and confidence intervals from mixed linear regressions with practice site modeled as a random effect.

Abbreviations: GAD-7, Generalized Anxiety Disorder-7; METs, Metabolic Equivalent of Task; PHQ-9, The Patient Health Questionnaire-9; PROMIS-29, Patient-Reported Outcomes Measurement Information System-29.

Figure 2. Effect of Practice Integration Profile (PIP) > 65 at baseline.



Notes: Each dot represents the adjusted difference in the outcome between patients in high- and low-integration practices, with 95% confidence intervals. *Higher scores indicate improved outcomes. ER = emergency room; CARE = Consultation and Relational Empathy scale

Limitations

The data we used may not generalize to all primary care practices and patients. The IBH-PC study enrolled only patients with multiple chronic medical and behavioral conditions; the results may not apply to patients with less complex presentations. All participating practices had at least some degree of IBH in that they had a BHC on site. It is unclear what effect, if any, stems from this minimum degree of IBH. All the practices were participating in a large-scale clinical trial designed to increase IBH, demonstrating commitment to this general model of care. Although they were similar to other practices around the nation in many measured respects, it is not clear if the effects of integration are similar in other settings.

Although the high- and low-integration practices were similar in most regards, their patients differed in many characteristics, such as markers of social and economic deprivation (Table 1). Therefore, we adjusted for these factors in multivariate regression. While the adjustment partially attenuated the effect of integration level, significant effects in several domains remained (Table 3). Additional unmeasured confounders may be unaccounted for.

We made many comparisons (19 at each time point), raising the possibility of false significance due to highlighting a small number of positive outcomes among a large number of tests. In this case, all 19 outcome measures were better in the high-integration practices, suggesting that random error is unlikely to explain the differences seen. If we were to apply the Bonferroni correction (dividing the nominal P of 0.05 by 19 = 0.00263)[51], at baseline only empathy would achieve significance. We note that there are arguments against adjusting for multiple comparisons.[52,53]

Baseline data were collected before the COVID-19 pandemic whereas follow-up data were collected until December 2020, when clinical care was significantly disrupted. This may have influenced the follow-up results, but not the baseline cross-sectional analyses.

The study used PIP 1.0. Since then, version 2 has been released (www.practiceintegrationprofile.com). Although it attempts to measure the same underlying constructs, many of the items were revised.[54] However, it has not yet been validated or used in a population with functional status measures as required for the analyses presented here. It is not clear how closely the 2 versions correlate or what an equivalent threshold score might be.

The threshold score of 65 was derived from the same study that was used to evaluate it. Although the threshold works well with many measures

beyond those used to derive it, including those collected 2 years later, it is not clear how it will perform in fully independent populations.

Conclusions

Patients in primary care practices with Total PIP scores of 65 or more reported better functional status, with significant differences in pain intensity, patient centeredness, empathy, social participation, physical function, the Physical Health Summary and the Duke Activity Status Index. Some, but not all, of this effect may be attributable to differing patient characteristics. A measurable target for Integrated Behavioral Health, such as Total PIP > 65, provides a focus for practice leadership and guidance on the time and resources needed to achieve a level of behavioral health service integration associated with desired outcomes for patients. This analysis provides further evidence of the broad, beneficial impacts of integrating behavioral health and primary care services. Future replication of this study with independent data sets will help confirm or disconfirm the relationship between PIP and patient outcomes reported here.

References

1. Kessler R, Stafford D. Primary Care IS the De Facto Mental Health System. In: Kessler R, Stafford D, eds. *Collaborative medicine case studies: evidence in practice* New York: Springer; 2008.
2. Auxier A, Runyan C, Mullin D, Mendenhall T, Young J, Kessler R. Behavioral health referrals and treatment initiation rates in integrated primary care: a Collaborative Care Research Network study. Behav Med Pract Policy Res 2012;2:337–44.
3. Butler M, Kane RL, McAlpine D, et al. Integration of mental health/substance abuse and primary care. Evid Rep Technol Assess (Full Rep) 2008;1–362.
4. Katon WJ, Lin EH, Von Korff M, et al. Collaborative care for patients with depression and chronic illnesses. N Engl J Med 2010;363:2611–20.
5. Miller BF, Kessler R, Peek CJ, Kallenberg GA. *A National Agenda for Research in Collaborative Care: Papers From the Collaborative Care Research Network Research Development Conference* Rockville, MD: Agency for Healthcare Research and Quality; 2011. AHRQ Publication No. 11-0067.
6. Baird M, Blount A, Brungardt S, Working Party Group on Integrated Behavioral Healthcare, et al. Joint principles: integrating behavioral health care into the patient-centered medical home. Ann Fam Med 2014;12:183–5.
7. Maeng DD, Poleshuck E, Rosenberg T, et al. Primary care behavioral health integration and care utilization: implications for patient outcome and healthcare resource use. J Gen Intern Med 2022;37:2691–7.
8. Thapa BB, Laws MB, Galarraga O. Evaluating the impact of integrated behavioral health intervention: evidence from Rhode Island. Medicine (Baltimore) 2021;100:e27066.
9. Jacobs C, Brieler JA, Salas J, Betancourt RM, Cronholm PF. Integrated behavioral health care in family medicine residencies: a CERA survey. Fam Med 2018;50:380–4.
10. Thomas H, Mitchell G, Rich J, Best M. Definition of whole person care in general practice in the English language literature: a systematic review. BMJ Open 2018;8:e023758.
11. Tong ST, Morgan ZJ, Stephens KA, Bazemore A, Peterson LE. Characteristics of family physicians practicing collaboratively with behavioral health professionals. Ann Fam Med 2023;21:157–60.
12. Filippi MK, Waxmonsky JA, Williams MD, et al. Integrated behavioral health implementation and training in primary care: a practice-based research network study. J Am Board Fam Med 2023;36:1008–19.
13. Hitt JR, Brennhofer SA, Martin MP, et al. Further experience with the practice integration profile: a measure of behavioral health and primary care integration. J Clin Psychol Med Settings 2022;29:274–84.
14. Dickinson WP. Strategies to support the integration of behavioral health and primary care: what have we learned thus far? J Am Board Fam Med 2015;28 Suppl 1:S102–106.
15. Peek CJ and the National Integration Academy Council. Lexicon Introduction. 2023; Available at: https://integrationacademy.ahrq.gov/products/ibh-lexicon. Accessed December 11, 2024.
16. Cohen DJ, Balasubramanian BA, Davis M, et al. Understanding care integration from the ground up: five organizing constructs that shape integrated practices. J Am Board Fam Med 2015;28:S7–S20.
17. Kessler R, Auxier A, Hitt JR, et al. Development and Validation of a Measure of Primary Care Behavioral Health Integration North American Primary Care Research Group Annual Meeting; November 14, 2016; Colorado Springs, CO.
18. Mullin DJ, Hargreaves L, Auxier A, et al. Measuring the integration of primary care and behavioral health services. Health Serv Res 2019;54:379–89.
19. Littenberg B, Clifton J, Crocker AM, et al. A cluster randomized trial of primary care practice redesign to integrate behavioral health for those who need it most: patients with multiple chronic conditions. Ann Fam Med 2023;21:483–95.
20. Crocker AM, Kessler R, van Eeghen C, et al. Integrating behavioral health and primary care

(IBH-PC) to improve patient-centered outcomes in adults with multiple chronic medical and behavioral health conditions: study protocol for a pragmatic cluster-randomized control trial. Trials 2021;22:200.
21. Integrating behavioral health and primary care (IBHPC) for comorbid behavioral and medical problems. Harvard Dataverse 2022; Available at: https://doi.org/10.7910/DVN/CT9PY6.
22. Cella D, Riley W, Stone A, et al. Initial adult health item banks and first wave testing of the Patient-Reported Outcomes Measurement Information System (PROMIS™) network: 2005–2008. J Clin Epidemiol 2010;63:1179–94.
23. Hays RD, Bjorner JB, Revicki DA, Spritzer KL, Cella D. Development of physical and mental health summary scores from the patient-reported outcomes measurement information system (PROMIS) global items. Qual Life Res 2009;18:873–80.
24. Hays RD, Spritzer KL, Fries JF, Krishnan E. Responsiveness and minimally important difference for the Patient-Reported Outcomes Measurement Information System (PROMIS) 20-item physical functioning short form in a prospective observational study of rheumatoid arthritis. Ann Rheum Dis 2015;74:104–7.
25. Khutok K, Janwantanakul P, Jensen MP, Kanlayanaphotporn R. Responsiveness of the PROMIS-29 scales in individuals with chronic low back pain. Spine (Phila Pa 1976) 2021;46:107–13.
26. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001;16:606–13.
27. Spitzer RL, Kroenke K, Williams JB, Lowe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006;166:1092–7.
28. Rose G, Bonnell L, Littenberg B, et al. Development of the Patient-Centeredness Index (PCI). North American Primary Care Research Group; November 20–24, 2020; Online.
29. Mercer SW, Maxwell M, Heaney D, Watt GC. The consultation and relational empathy (CARE) measure: development and preliminary validation and reliability of an empathy-based consultation process measure. Fam Pract 2004;21:699–705.
30. Hlatky MA, Boineau RE, Higginbotham MB, et al. A brief self-administered questionnaire to determine functional capacity (the Duke Activity Status Index). Am J Cardiol 1989;64:651–4.
31. Schwark EH. Relationship of fitness and body mass index to cardiac disease. JAMA 2005;293:161–2; author reply 162.
32. MacLean CD, Littenberg B, Gagnon M, Reardon M, Turner PD, Jordan C. The Vermont Diabetes Information System (VDIS): study design and subject recruitment for a cluster randomized trial of a decision support system in a regional sample of primary care practices. Clin Trials 2004;1:532–44.
33. Adams PF, Hendershot GE, Marano MA. Current estimates from the National Health Interview Survey, 1996. Vital Health Stat 1999;10:141.
34. Macchi CR, Kessler R, Auxier A, et al. The practice integration profile: rationale, development, method, and research. Fam Syst Health 2016;34:334–41.
35. Kessler RS, Auxier A, Hitt JR, et al. Development and validation of a measure of primary care behavioral health integration. Fam Syst Health 2016;34:342–56.
36. Cleveland WS. Robust locally weighted regression and smoothing scatterplots. Journal of the American Statistical Association 1979;74:829–36.
37. Mickey RM, Greenland S. The impact of confounder selection criteria on effect estimation. Am J Epidemiol 1989;129:125–37.
38. Greenland S. Modeling and variable selection in epidemiologic analysis. Am J Public Health 1989;79:340–9.
39. Totura CMW, Fields SA, Karver MS. The role of the therapeutic relationship in psychopharmacological treatment outcomes: a meta-analytic review. Psychiatr Serv 2018;69:41–7.
40. Hausman A. Modeling the patient-physician service encounter: improving patient outcomes. Journal of the Academy of Marketing Science 2004;32:403–17.
41. Novikov Z, Glover WJ, Trepman PC, Naveh E, Goldfracht M. How do integrative practices influence patient-centered care? An exploratory study comparing diabetes and mental health care. Health Care Manage Rev 2016;41:113–26.
42. Peikes D, Genevro JL, Scholle SH, Torda P. The patient-centered medical homes strategies to put patients at the center of primary care (issue brief). Mathematica Policy Research Reports 2011.
43. Karazivan P, Dumez V, Flora L, et al. The patient-as-partner approach in health care: a conceptual framework for a necessary transition. Academic Medicine 2015;90:437–41.
44. Peek CJ and the National Integration Academy Council. *Lexicon for Behavioral Health and Primary Care Integration* Rockville, MD: Agency for Healthcare Research and Quality; 2013. AHRQ Publication No. 13-IP001-EF.
45. Cubillos L, Bartels SM, Torrey WC, et al. The effectiveness and cost-effectiveness of integrating mental health services in primary care in low- and middle-income countries: systematic review. BJPsych Bull 2021;45:40–52.
46. Huffman JC, Niazi SK, Rundell JR, Sharpe M, Katon WJ. Essential articles on collaborative care models for the treatment of psychiatric disorders in

medical settings: a publication by the academy of psychosomatic medicine research and evidence-based practice committee. Psychosomatics 2014;55:109–22.

47. Rodgers RC, Hunter JE. Impact of management by objectives on organizational productivity. Journal of Applied Psychology 1991;76:322–36.
48. Williams NJ, Wolk CB, Becker-Haimes EM, Beidas RS. Testing a theory of strategic implementation leadership, implementation climate, and clinicians' use of evidence-based practice: a 5-year panel analysis. Implement Sci 2020;15:10.
49. Birken S, Clary A, Tabriz AA, et al. Middle managers' role in implementing evidence-based practices in healthcare: a systematic review. Implement Sci 2018;13:149.
50. Sampath BRJ, Baldoza K, Mate K, Lenoci-Edwards J, Barker P. *Whole System Quality: A Unified Approach to Building Responsive, Resilient Health Care Systems* Boston, MA: Institute for Healthcare Improvement; 2021.
51. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. BMJ 1995;310:170.
52. Rothman KJ. No adjustments are needed for multiple comparisons. Epidemiology 1990;1:43–6.
53. Feise RJ. Do multiple outcome measures require p-value adjustment? BMC Med Res Methodol 2002;2:8.
54. Rose GL, Weldon TL, McEntee ML, et al. Practice integration profile revised: improving item readability and completion. Fam Syst Health 2023;41:201–6.